
Günter Kochendörfer

Kortikale Linguistik

Teil 2: Grundlagen



<http://www.cortical-linguistics.de>
15. 3. 2004

Gesamtinhaltsübersicht

Teil 1: Wissenschaftstheoretische Voraussetzungen

Teil 2: Grundlagen

Teil 3: Phonetik/Phonologie

Teil 4: Lexikon, Morphologie

Teil 5: Syntax

Teil 6: Gedächtnisformen, Textverstehen

Teil 7: Denken und Formulieren

Teil 8: Spracherwerb

Teil 9: Sprachpathologie

Teil 10: Randgebiete

Anhang: Software

Literatur

Index

Teil 2

Grundlagen

Der Kortex kann als universelles datenverarbeitendes System verstanden werden und Sprache kann in einem solchen System beliebig hardwarefern strukturiert sein (ohne je hardwareunabhängig zu werden). Unter der Annahme großer Hardwareferne wäre ein Blick auf kortikale Strukturen für die Linguistik eher nebensächlich. Eine wissenschaftlich fruchtbarere Hypothese ist allerdings, dass sich Sprache in der Evolution in möglichst großer Nähe zu bereits vorhandenen kortikalen Strukturen entwickelt hat, das heißt unter optimaler Verwendung kortikaler Strukturprinzipien, die auch in anderen Leistungsdomänen gültig sind.

Wenn man die Hypothese der Hardwarenähe von Sprache zugrunde legt, stößt man zunächst ganz allgemein auf das Problem, wie irgendwelche Inhalte, also nicht nur sprachliche, im Kortex repräsentiert zu denken sind. Die Meinungen darüber gehen sehr weit auseinander. Von der Linguistik her kommend kann man Berührungspunkte erkennen zwischen dem allgemeinen Repräsentationsproblem und zentralen Konzepten der linguistischen Semantik, insbesondere der Prototypentheorie.

Aus den Besonderheiten der Datenverarbeitung im Kortex ergibt sich, dass die Repräsentation von Inhalten und die dazugehörigen Lernprozesse immer zusammen mit Produktions- und Perzeptionsprozessen zu behandeln sind. Das heißt, dass gerade auch auf das Verständnis der Produktionsprozesse (Top-down-Zugriff auf Repräsentationen, Entstehung von Vorstellungen usw.) besonderes Gewicht gelegt werden muss.

Inhalt

2.1 Biologische Grundannahmen	9
2.1.1 Nervensystem und Gehirn	
2.1.2 Der Kortex	
2.1.3 Das Neuron	
2.1.4 Simulation neuronaler Grundfunktionen	
2.2 Die Repräsentationsproblematik	31
2.2.1 Allgemeine Bedingungen für mentale Repräsentationen	
2.2.2 Symbolverarbeitende Modelle	
2.2.3 Konnektionismus	
2.2.4 Zellenassemblies	
2.2.5 Zusammenfassende Bewertung	
2.3 Lokalistische Repräsentation von Konzepten	55
2.3.1 Beobachtungen	
2.3.2 Lokalistische Repräsentation, Lernprozesse und Einzelimpulskodierung	
2.3.3 Konzeptbildung	
2.3.4 Konsequenzen für das Verständnis kortikaler Strukturen	
2.3.5 Diskussion	
2.4 Semantik	83
2.4.1 Fragestellung	
2.4.2 Bedeutungsbeschreibung und Realitätsbezug	
2.4.3 Kategorisierung	
2.4.4 Wortfelder, Bedeutungsverwandtschaft und distinktive Merkmale	

2.4.5	Prototypizität	
2.4.6	Kontexte	
2.4.7	Szenen	
2.4.8	Abschließendes Experiment zur Prototypentheorie	
2.5	Vorstellungen	127
2.5.1	Produktion von Vorstellungen als Umkehrung der Perzeption	
2.5.2	Rückspiegelung	
2.5.3	Neuronale Grundlagen und Lernvorgänge	
2.5.4	Kreative Vorstellungsprozesse	
2.6	„Adressierung“ und „Zugriff“ im Kortex	153
2.6.1	Adressierung	
2.6.2	Zugriff	
2.6.3	Konsequenzen	

2.1 Biologische Grundannahmen

Das Kapitel 2.1 ist nicht als Einführung in die Neurobiologie gedacht, sondern enthält Hinweise auf Elemente, die in den folgenden Kapiteln und innerhalb des Gesamtunternehmens „kortikale Linguistik“ bestimmend sind. Die Darstellung orientiert sich an gängigen Handbüchern: Schmidt, Thews & Lang (2000), Schmidt & Schaible (2001), Kandel, Schwartz & Jessell (1996), Klinke & Silbernaagl (2003). Fundstellen in diesen Handbüchern werden nur in speziellen Fällen bzw. bei direkten Zitaten nachgewiesen. Die Darstellung wendet sich in erster Linie an Linguisten, weniger an Biologen oder Neurologen. Sie enthält allgemein akzeptierte Informationen, nicht solche Annahmen über biologische Strukturen und Prozesse, die sich später speziell aus der Entwicklung sprachverarbeitender Modelle ergeben.

2.1.1 Nervensystem und Gehirn

Das Stichwort „kortikale Linguistik“ lässt an eine Linguistik denken, die sich speziell oder vielleicht sogar ausschließlich mit Sprache im Kortex beschäftigt. Ein solches Unternehmen wäre aber letztlich sinnlos. Man muss sich von vornherein klar machen, dass die Gesamtheit der am Phänomen Sprache beteiligten Instanzen zu beachten ist, und das bedeutet nicht nur das Zentralnervensystem oder das Gehirn mit seinen verschiedenen Abschnitten oder gar nur der Kortex, sondern das Gehirn einschließlich der Sinnesperipherie bzw. motorischen Peripherie und der Verbindungen vom und zum Kortex. Das ergibt sich schon allein aus den modelltheoretischen Forderungen der Präzisierbarkeit und Einbettbarkeit eines als korrekt zu akzeptierenden Modells (vgl. Teil 1). In Kapitel 2.3.3 wird darüber hinaus gezeigt, dass ein Gehirn ohne Sinnesperipherie und/oder Motorik keine Information enthält, also die informationsverarbeitende Leistung des Gehirns nur ver-

standen werden kann, wenn Gehirn und peripheres Nervensystem als Einheit betrachtet werden.

Selbst wenn man sich nur auf das Nervensystem ohne Einschluss der Sinnesorgane und der Motorik beschränkt, kann man an einer Einteilung wie der folgenden erkennen, dass der Kortex nur eine unter einer ganzen Reihe von Komponenten ist:

Auto- nomes Nerven- system	Zerebrospinales Nervensystem				
	Rücken- mark	Gehirn			
		Myelen- cephalon = Medulla oblongata	Meten- cephalon mit z. B. dem Kleinhirn	Mesen- cephalon	Dien- cephalon mit z. B. dem Thalamus

Auch was speziellere Fragestellungen angeht, ist die Beachtung des Zusammenspiels von Kortex und Peripherie eine ganz wesentliche Informationsquelle. Ein klassisches Beispiel findet sich bei Lenneberg (1972). Lenneberg stellt über mehrere Buchseiten hinweg Überlegungen darüber an, was sich aus den neuronalen Verbindungen, speziell solchen zur Motorik des Kehlkopfs und den Artikulationsorganen des Mundraums, für die Verarbeitungsprozesse in der Lautproduktion ergibt.

Er geht aus von Informationen über Durchmesser und Länge beteiligter Nervenfasern, wie in folgender Tabelle (Lenneberg, 1972: 123, etwas verkürzt) wiedergegeben:

Hirnnerv	Länge in cm	Faserdurchmesser in μm (Mittelwert)
Zweig zum inneren Flügelmuskel (Kiefer)	9,7	9,5
Zweig zur Oberlippe	26,4	10,3
rückläufiger Nerv (Kehlkopf)	31,7	5,4

Das heißt, gerade die längste der an den betrachteten Prozessen beteiligten Fasern ist gleichzeitig auch die dünnste. Diese Daten, zusammen mit der Tatsache, dass die Geschwindigkeit der Signalweiterleitung bei dünnen Fasern geringer ist als bei dicken, führen zur Feststellung (ebenda)

„... daß die Anatomie der Nerven darauf hindeutet, daß die Innervationszeit für die Muskeln innerhalb des Kehlkopfes leicht bis zu 30 msec länger sein kann als die Innervationszeit für die Muskeln in und um die Mundhöhle.

Wenn wir nun bedenken, daß einige Artikulationsereignisse nicht länger als 20 msec dauern, erscheint die Annahme sinnvoll, daß die Auslösungsordnung im Hirnstamm manchmal von der Ordnung der Ereignisse an der Peripherie verschieden ist.“

Aus dieser Überlegung folgt schließlich die Ablehnung einer kettenförmigen Modellierung der Lautproduktion, bei der die auf den Reiz einer Muskelgruppe folgende Reaktion den nächsten Reiz auslöst. Die Konsequenzen für das Verständnis der Phonologie sind sehr einschneidend.

Da das Argument auf Annahmen über die Ausbreitungsgeschwindigkeit neuronaler Signale auf Nervenfasern beruht, lohnt allerdings ein Blick in neuere Lehrbücher. Dort liest man Zahlen, die die Lenneberg'sche Schlußfolgerung relativieren. Hier ein Ausschnitt aus einer Tabelle bei Schmidt & Schaible (2001: 49):

Fasertyp, Funktion	Mittlerer Faser- durchmesser [μm]	Mittlere Leitungs- geschwindigkeit [m/s]
Primäre Muskelspindelafferenzen motorisch zu Skelettmuskeln	15	100(70-120)
Hautafferenzen für Berührung und Druck	8	50(30-70)
Motorisch zu Muskelspindeln	5	20(15-30)

Wenn wir grob die durchschnittliche Geschwindigkeit der Potentialausbreitung auf den betroffenen Fasern mit 50m/s ansetzen, gelangen wir problemlos in einen Bereich, der mit der Dauer von Kurzvokalen bei Annahme von Top-down-Aktivierung und anschließender Rückmeldung vereinbar ist. Die von Lenneberg ausgerechnete Differenz von 30ms reduziert sich auf vielleicht 10ms oder weniger. Lenneberg ist also, vom heutigen Standpunkt aus gesehen, möglicherweise von zu geringen Leitungsgeschwindigkeiten ausgegangen. Die Lehre, die auch heute noch Bestand hat, ist aber, dass die Einbeziehung der Peripherie in die linguistischen Überlegungen auch dann erforderlich ist, wenn es nicht speziell um motorische oder sensorische Phänomene geht.

Andererseits ist es aber auch so, dass die direkten Beziehungen zur Peripherie nur einen kleinen Teil der mentalen Leistungen erklären oder bedingen können. Das ergibt sich schon allein daraus, dass die primären Sinnesareale nur einen Bruchteil der Kortexoberfläche ausmachen (Abbildung 2.1.1-1).

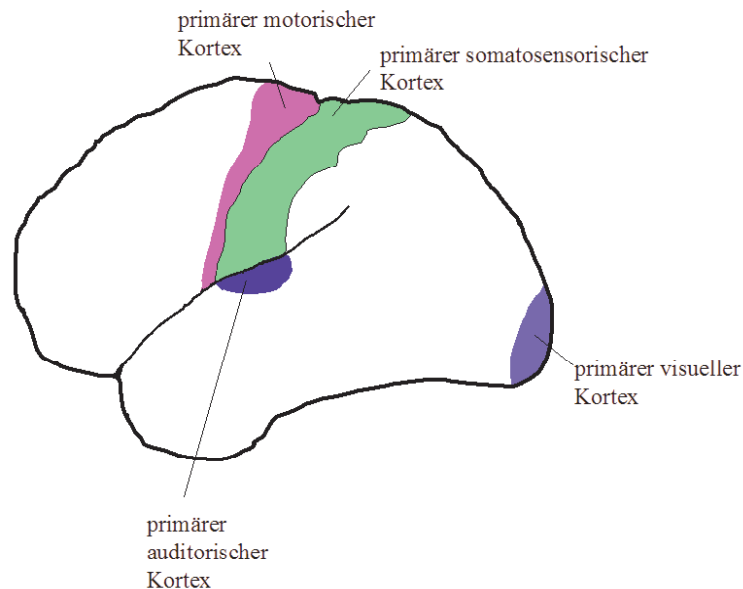


Abbildung 2.1.1-1: Umriss des Kortex (linke Hemisphäre) mit Andeutung der „primären“ Areale.

Das muss Konsequenzen für das Verständnis der Verarbeitungsprozesse generell und nicht nur im sprachlichen Bereich haben, z. B. für die Ausein-

andersetzung mit konnektionistischen Modellen, vgl. unten Kapitel 2.2. Offenbar ist entweder die Umsetzung von Reizen zu Reaktionen nicht im wesentlichen direkt (was immer das heißen mag), oder aber es gibt Strukturen und Prozesse, die in irgendeinem Sinne „oberhalb“ des Reiz-Reaktions-Zusammenhangs angesiedelt sind.

Man muss also grundsätzlich mit einer bestimmenden Rolle spezifisch kortikaler Strukturen rechnen, und es ist vielleicht auch schon von vornherein wahrscheinlich oder wenigstens als mögliche Annahme plausibel, davon auszugehen, dass solche Strukturen – wie allgemein sie auch immer sein mögen – einen Einfluss auf die Struktur natürlicher Sprachen haben werden.

Das Attribut „kortikal“ in der Bezeichnung „kortikale Linguistik“ meint also nicht eine Beschränkung auf den Bereich des *cortex cerebri*, sondern bezeichnet eher eine Perspektive. Dabei wird aber kortikalen Strukturen sozusagen eine Rolle im Vordergrund zugewiesen.

2.1.2 Der Kortex

Die folgende Definition von Braitenberg & Schüz (1998: 7) stützt sich hauptsächlich auf bestimmte strukturelle Merkmale:

„Tentatively, we define the cortex as that part of the grey substance in the anterior part of the brain which has a planar layout, with a fairly homogeneous appearance in every direction of the plane but with an ordered structure at right angles to it, the so-called layers.“

Der Kortex (*cortex cerebri*) hat eine Fläche von ungefähr 1000 cm^2 und ist nur ungefähr zwei bis vier Millimeter dick. Die Zahl der kortikalen Neuronen wird mit ungefähr 10^{10} angegeben. Daneben enthält der Kortex eine noch wesentlich größere Zahl von sog. „Gliazellen“ unterschiedlicher Typen, die keine datenverarbeitende Funktion haben.

Die Schichtung, von der in dem Zitat aus Braitenberg & Schüz die Rede war, ist schematisch in Abbildung 2.1.2–1 wiedergegeben.

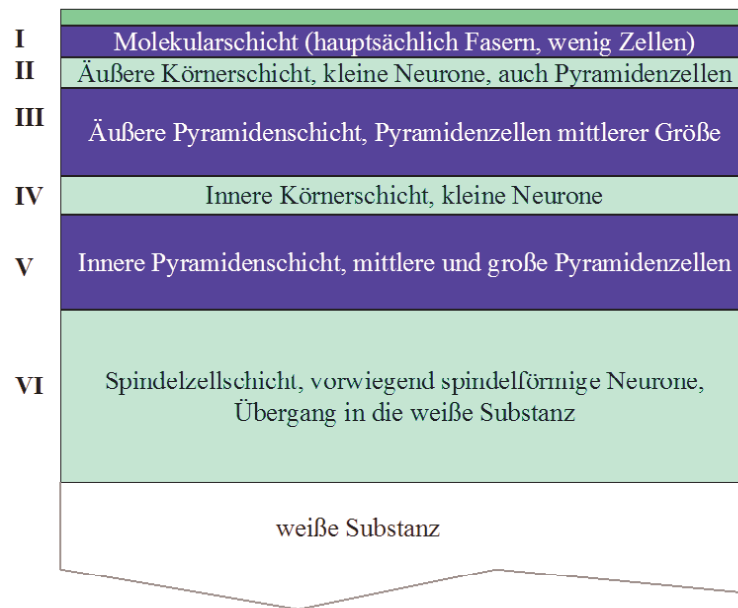


Abbildung 2.1.2–1: Einfaches Schema der Schichtung des Kortex. Die Schichtung ist nicht in allen Bereichen des Kortex gleich.

Es ist zu beachten, dass der Eindruck der Schichtung nicht dadurch entsteht, dass sich gleichartige Strukturen in schichtartiger Abgrenzung wiederholen, sondern die einzelnen Schichten werden dadurch erkennbar, dass sie unterschiedliche Zelltypen (oder Zellen unterschiedlicher Größe bzw. Form) enthalten oder relativ frei von Zellkörpern sind.

Wenn man den Überlegungen von Braitenberg & Schüz (1998: 67 ff.) folgt, genügt es, drei Zelltypen zu unterscheiden: Pyramidenzellen, Sternzellen und Martinotti-Zellen. Dabei überwiegen die Pyramidenzellen mit 85%, der Rest verteilt sich auf Sternzellen und Martinotti-Zellen, deren Anteile schwer zu bestimmen sind. Die Zelltypen werden nach ihrer Form unterschieden. Ein Beispiel für eine Pyramidenzelle wird unten in Abbildung 2.1.3–1 gegeben. Die Sternzellen der Klassifikation von Braitenberg & Schüz sind dadurch charakterisiert, dass sie keine dendritischen Dornen („spines“) haben, die Martinotti-Zellen dadurch, dass das Axon (zum Begriff siehe den folgenden Abschnitt) nicht, wie bei Pyramidenzellen, im wesentlichen in absteigender, sondern in aufsteigender Richtung verläuft.

Der Kortex zeigt ein unglaublich dichtes, chaotisch erscheinendes Verbindungsgeflecht. Man kann nach ihrem Verlauf verschiedene Typen von Verbindungen unterscheiden, wie in Abbildung 2.1.2–2 zusammengestellt.

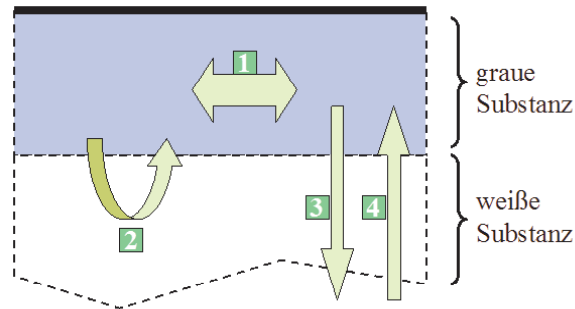


Abbildung 2.1.2–2: Typen von Verbindungen. **1.** Lokale Verbindungen innerhalb des Kortex. **2.** Kortiko-kortikale Verbindungen über die weiße Substanz. **3.** Verbindungen, die den Kortex verlassen. **4.** Verbindungen, die in den Kortex hineinführen. Weitere Erläuterungen im Text.

Die meisten Verbindungen bleiben lokal innerhalb des Kortex (der grauen Substanz). Sie werden gebildet durch waagrechte Verzweigungen der Axone von Pyramidenzellen sowie durch die Axone von Sternzellen und Martinotti-Zellen.

Die Verbindungen, die den Kortex verlassen oder in ihn hineinführen, bilden die weiße Substanz. Darunter sind mehr als 90% kortiko-kortikale Fasern, das heißt Fasern, die Verbindungen mit anderen Stellen des Kortex bilden. Man kann das als Hinweis darauf nehmen, dass Verarbeitungsergebnisse, also Output des Kortex, in großem Umfang wieder als Input in den Kortex dienen und dort weiterverarbeitet werden.

Die übrigen Fasern bilden Verbindungen z. B. zur Motorik oder von den Sinnesorganen.

Fasern in der weißen Substanz sind typischerweise myelinisiert, das heißt mit einer von Gliazellen gebildeten Hülle versehen, die bewirkt, dass die Reizweiterleitung wesentlich beschleunigt wird.

Verschiedene Bereiche des Kortex sind auf verschiedene Aufgaben spezialisiert. Das ist ein Ergebnis klassischer Läsionsstudien ebenso wie moderner bildgebender Untersuchungen am arbeitenden Gehirn. Es ist aber offenbar so, dass damit nicht eine große strukturelle Differenzierung einher geht. Die Differenzen bestehen hauptsächlich darin, dass bestimmte Schichten

(vgl. Abbildung 2.1.2–1) mehr oder weniger ausgeprägt erscheinen, nicht dadurch, dass die neuronalen Verschaltungen im Detail völlig unterschiedlich wären (das gilt mit Ausnahme der besonderen Region des Hippokampus). Die relativ einheitlichen strukturellen Merkmale sollten, so ist zu vermuten, ein Äquivalent in einer relativ einheitlichen Funktion haben.

Schon allein aus der dichtgepackten Verbindungsstruktur ergibt sich, dass Lernprozesse, das heißt die Repräsentation neuer Inhalte und neuer Fähigkeiten, nicht darin bestehen können, dass Verbindungen in großem Umfang neu wachsen, geschweige denn, dass sie bei Vergessensprozessen einfach wieder verschwinden. Man kann sich auch nicht vorstellen, dass Zellen in einer informationsspezifischen Verschaltung (bei Erwachsenen und abhängig von Wahrnehmungsprozessen) neu entstehen. Wohl aber können neuronale Strukturen unabhängig von spezifischen Verarbeitungsprozessen abgebaut werden.

2.1.3 Das Neuron

Neuronen sind Zellen, die zwar viele Eigenschaften mit anderen Zellen gemeinsam haben, deren Besonderheit aber darin besteht, dass sie elektrische Signale weiterleiten können.

Man kann im wesentlichen drei Bereiche an einer solchen Zelle unterscheiden: den Zellkörper, das Axon und die Dendriten(bäume). Es gibt nur ein einziges verzweigtes Axon, aber ggf. viele Dendriten. Das Axon ist die Struktur, über die Signale an nachgeschaltete Zellen weitergegeben werden, die Dendriten sind die Eingangsstrukturen, über die Signale in die Zelle hineingelangen. Die Dendriten der Pyramidenzellen tragen zahlreiche sog. dendritische Dornen, denen vielleicht eine Bedeutung im Zusammenhang mit Lernprozessen zukommt. Die Enden der Axone haben keine anatomische Verbindung mit den Dendriten der nachgeschalteten Zellen, sondern sind durch einen Spalt davon getrennt. Die Signalübertragung geschieht durch Ausschüttung eines chemischen Neurotransmitters in diesen Spalt. Die so gebildeten Kontaktstellen heißen Synapsen. Von Braitenberg & Schüz (1998:107) wird vermutet, dass die Zahl der Synapsen auf dornenbesetzten Dendriten im Wesentlichen der Zahl der Dornen entspricht. Die Gesamtzahl der Synapsen auf Pyramidenzellen wird unterschiedlich angegeben, die Zahl 10000 stellt wohl einen eher niedrigen Schätzwert dar.

Die Abbildung 2.1.3–1 zeigt eine Pyramidenzelle als typisches Beispiel für die Gestalt und die wichtigsten morphologischen Eigenschaften eines Neurons.

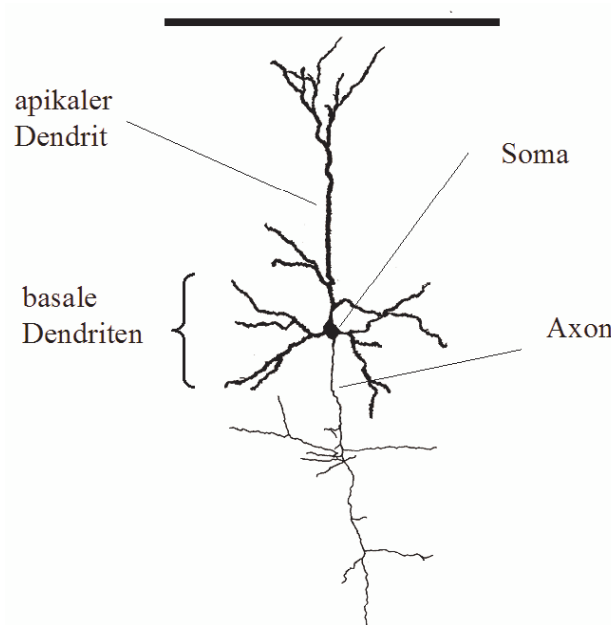


Abbildung 2.1.3-1: Neuron (Pyramidenzelle) in Golgi-Färbung, verändert nach Braitenberg & Schüz (1998: 70). Gesamthöhe der abgebildeten Zelle 0,5 mm. Die Dendriten erscheinen verdickt und unregelmäßig durch die zahlreichen dendritischen Dornen, mit denen sie besetzt sind. Der dicke waagrechte Strich deutet die Kortexoberfläche an.

Die Asymmetrie zwischen den vielen dendritischen Eingängen in die Zelle und dem einen Ausgang über das Axon entspricht einem wichtigen Verarbeitungscharakteristikum im Nervensystem und im Kortex. Das wird deutlich, wenn man in einem Gedankenexperiment eine neuronale Struktur konstruiert, die nicht aus biologisch realistischen Neuronen besteht, sondern aus fiktiven Neuronen, bei denen die Zahlenverhältnisse genau umgekehrt sind, die also einen Eingang und zahlreiche mit unterschiedlichen Signalen beschickbare Ausgänge haben. Der Unterschied ist in Abbildung 2.1.2-2 veranschaulicht.

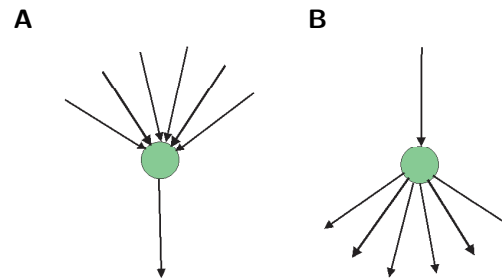


Abbildung 2.1.3-2: Gedankenexperiment zur Asymmetrie von Eingängen und Ausgängen am Neuron. **A.** Biologisches Neuron. **B.** Fiktives Neuron.

Strukturen, die ausschließlich aus dieser Art fiktiver Neuronen bestünden, müssten streng baumförmig sein. Für die Steuerung der Motorik könnte es von Vorteil sein, Nervenfasern zu verschiedenen Muskeln von einer einzelnen Zelle aus anzusteuern. Für die Wahrnehmungsseite ergeben sich aber entscheidende Merkwürdigkeiten. Da es nicht vorgesehen ist, dass verschiedene Wahrnehmungsdetails zu einem Ganzen integriert werden können, müssen alle Wahrnehmungskonzepte primitiv und angeboren sein. Ein solcher Organismus wäre unfähig zur Anpassung an eine sich wandelnde Umgebung.

Die fiktiven Neuronen des Gedankenexperiments sind also vor allem deshalb keine geeigneten Verarbeitungseinheiten, weil sie keine flexible Konzeptbildung in der Wahrnehmung erlauben. Für biologische Neuronen folgt daraus aber, dass sie offenbar in besonderer Weise wahrnehmungsorientiert strukturiert sind.

Für die nun folgende Beschreibung der Funktion von Neuronen ist zu beachten, dass Neuronen innerhalb des Projekts „Kortikale Linguistik“ durchweg als Leistungsmodelle abgebildet werden. Das heißt, dass ihre „innere“ Struktur als nicht detailliert gilt und nur Eigenschaften oder Verhaltensweisen, die auf die „äußere“ strukturelle und funktionale Umgebung Einfluss haben, mit modellhafter Gültigkeit festgehalten werden. Das bedeutet, dass neurophysiologische Informationen beliebiger Art zwar herangezogen werden, das Ziel aber in der Klärung der leistungsmodellhaften Charakterisierung der einzelnen Neuronen besteht. Alle in den Modellen behaupteten Strukturen oberhalb der Ebene des einzelnen Neurons haben dagegen – in den jeweils angegebenen oder implizierten Beschränkungen – die für Funktionsmodelle zu fordernde Gültigkeit. (Zur Unterscheidung von Leistungsmodell und Funktionsmodell siehe Teil 1.)

Damit kann ein Neuron als eine punktförmige Einheit betrachtet werden, die aufgrund eines Inputs Signale abgibt. Die abgegebenen Signale heißen

Aktionspotenziale. Sie haben eine Amplitude von ca. 100 mV. Aktionspotenziale können zwar unter bestimmten Umständen auch verschieden große Amplituden haben, es gibt aber keinerlei Anhaltspunkte dafür, dass damit Information kodiert werden könnte. Auch die Zeitdauer eines Aktionspotenzials, die größenordnungsmäßig mit 1 ms angegeben wird, ist nicht systematisch veränderbar.

Information kann über Aktionspotenziale grundsätzlich nur kodiert werden

- durch die *Auswahl des Axons*, auf dem das Aktionspotenzial weitergeleitet wird (labeled line coding, vgl. Kandel, Schwartz & Jessell, 1996: 24),
- durch eine *Folge* von Aktionspotenzialen (z. B. rate coding),
- durch ein *zeitliches Verhältnis* eines einzelnen Aktionspotenzials zu anderen Ereignissen.

Ein Aktionspotenzial wird getriggert, wenn das Membranpotenzial einer Nervenzelle einen Schwellenwert überschreitet, der in der Gegend von -50mV liegt. Das Membranpotenzial einer Zelle in Ruhe (Ruhepotenzial) liegt bei -70mV. Eine Verschiebung des Zellpotenzials in positiver Richtung wird durch ein exzitatorisches postsynaptisches Potenzial (EPSP), in negativer Richtung durch ein inhibitorisches postsynaptisches Potenzial (IPSP) bewirkt. EPSPs und IPSPs werden durch synaptische Aktivität von vorgeschalteten Zellen ausgelöst. Ihre Größe ist von der Effektivität der Synapsen (den Synapsengewichten), nicht von Signaleigenschaften der vorgeschalteten Zelle abhängig. Oft wird auch einfach das Membranpotenzial einer Zelle als Ganzes oberhalb des Ruhepotenzials als EPSP und unterhalb des Ruhepotenzials als IPSP bezeichnet. Wenn ein exzitatorisches Membranpotenzial die Schwelle zur Auslösung eines Aktionspotenzials nicht erreicht, klingt es exponentiell ab. Die Wirkungen, die kurz aufeinanderfolgende Erregungen der Zelle haben, addieren sich, sodass zwei oder mehrere für sich genommen unterschwellige, aber in kurzem Abstand auftretende EPSPs die Entstehung eines Aktionspotenzials auslösen können.

Das Feuern einer Zelle (die Bildung eines Aktionspotenzials) hat entweder erregende oder hemmende Wirkung auf eine nachgeschaltete Zelle. Es gibt keine Zellen, die sowohl erregende als auch hemmende Synapsen als Ausgänge bilden. Auf einer nachgeschalteten Zelle können aber sowohl erregende als auch hemmende Synapsen gebildet werden.

Die typischen Verläufe eines unterschwelligen EPSP und eines Aktionspotenzials sind in Abbildung 2.1.3–3 dargestellt.

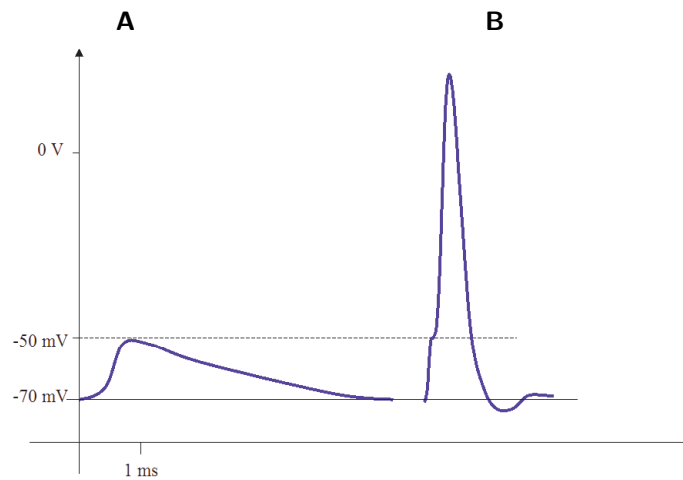


Abbildung 2.1.3-3: Signalverläufe an einer Nervenzelle. **A.** Unterschwelliges EPSP. **B.** Aktionspotenzial.

In der Abbildung kann man erkennen, dass auf das Aktionspotenzial folgend das Membranpotenzial nicht sofort auf das Ruhepotenzial zurückgeht, sondern eine kurze Zeit unterhalb des Ruhepotenzials verläuft. In dieser Zeit ist die Zelle unerregbar. Diese als Refraktärphase bezeichnete Zeitspanne liegt, wie die Dauer des Aktionspotenzials, im Millisekundenbereich.

Wichtig ist, dass die gradierten EPSPs und IPSPs lokale Phänomene innerhalb des Neurons sind und nicht weitergeleitet werden. Das weitergeleitete Signal, also das Aktionspotenzial, ist ein Alles-oder-nichts-Signal und zwar in der Weise, dass es zusätzlich in seiner Länge nicht variierbar ist. Signalverläufe entsprechend der Abbildung 2.1.3-4, wie man sie in der digitalen Elektronik antrifft, sind im Nervensystem nicht möglich.

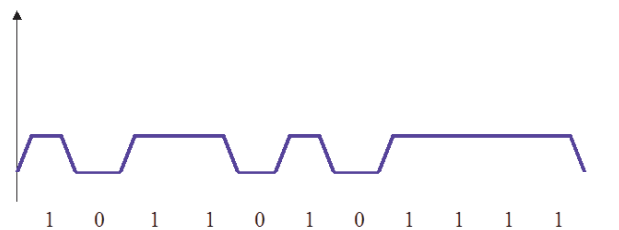


Abbildung 2.1.3-4: Beispiel eines technischen Signalverlaufs, z. B. auf einer Datenübertragungsstrecke.

Alle Veränderungen von Potenzialen, ihre Ausbreitung innerhalb der Zelle und die Weiterleitung des Aktionspotenzials sind Vorgänge, die Zeit brauchen. Wenn man die Zelle als punktförmig betrachtet und sie damit sozusagen auf den Ursprung des Axons, den sog. Axonhügel, der die Triggerzone für die Auslösung des Aktionspotenzials ist, schrumpft, muss man die Ausbreitungszeiten innerhalb der Zelle dem Zeitverbrauch auf den Verbindungen zwischen den Zellen zuschlagen. Da der auf diese Weise zustandekommende Zeitbedarf von Verbindung zu Verbindung etwas verschieden sein wird, muss man mit einer entsprechenden Beeinflussung der Integration der Eingangssignale in der Zelle rechnen.

Das Zustandekommen des Aktionspotenzials und die Tatsache, dass alle anderen Zustände einer Nervenzelle lokal bleiben, ist unvereinbar mit der Vorstellung, dass Informationen, die in neuronale Verarbeitungsprozesse einbezogen werden sollen, im *Innern* der Zelle gespeichert sind. Die Adressierung und Weiterleitung solcher Informationen über einen aus Aktionspotenzialen bestehenden Kode müsste völlig unerklärt bleiben. Zusammen mit der oben in Abschnitt 2.1.2 angesprochenen Unmöglichkeit, Verbindungen nach Bedarf neu entstehend zu denken, ergibt sich damit als plausible Möglichkeit für Speicherung und Vergessen nur das Verstärken oder Abschwächen bereits vorhandener Verbindungen, was identisch ist mit der Verstärkung oder Abschwächung von Synapsen. Es gibt Anhaltspunkte dafür, dass davon nur exzitatorische, nicht auch inhibitorische Synapsen betroffen sein können (Braitenberg & Schüz, 1998:181).

2.1.4 Simulation neuronaler Grundfunktionen

Eine genauere Beschreibung des verwendeten Simulationsprogramms und begleitender Software wird in einem eigenen Teil des Gesamtprojekts gegeben (Anhang: Software). Hier werden zunächst nur einige Informationen zusammengestellt, die für das Verständnis der folgenden Simulationsbeispiele nützlich sind.

Es wird eine Simulationstechnik verwendet, bei der das Zeitkontinuum in Zeittakte aufgelöst ist und innerhalb eines Zeittakts alle sich verändernden Werte, die ein zu simulierendes Netz charakterisieren, neu berechnet werden. Innerhalb eines Zeittakts steht die Zeit sozusagen still.

Das Simulationsprogramm realisiert ausschließlich das Verhalten von Zellen, es gibt keine übergeordneten Prozesse (Monitoring-Prozesse, Lernalgorithmen und dergleichen). Zellen können über eine Eingabedatei zu bestimmten Zeitpunkten mit einem externen Input (einem überschwelligem EPSP)

versehen werden, womit ein Ersatz für die Sinneswahrnehmung oder für Ergebnisse nicht modellierter begleitender Verarbeitungsprozesse geschaffen wird.

Die Abbildung 2.1.4–1 gibt die verwendeten Zellparameter wieder. Es muss unterschieden werden zwischen Parametern, die für Zellen insgesamt gelten und Parametern, die für einzelne Verbindungen zu spezifizieren sind. Einige der Parameter sind durch Verarbeitungsprozesse veränderlich, andere sind konstant.

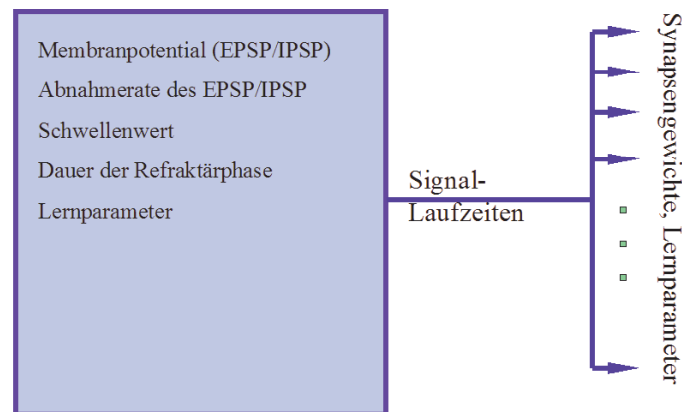


Abbildung 2.1.4–1: Die Spezifizierung des Modellneurons (übernommen aus Kochendörfer, 2002: 20).

Zu den einzelnen Parametern (verkürzt nach Kochendörfer, 2002: 21):

Das *Membranpotential* unterliegt, solange es nicht identisch ist mit dem Ruhepotential der Zelle, ständiger Veränderung. Es ist in Simulationen üblich, das Ruhepotential mit 0 anzunehmen und die exzitatorischen und inhibitorischen postsynaptischen Potentiale entsprechend darauf abzustimmen. Die Geschwindigkeit, mit der sich ein nicht durch Input verändertes Potenzial dem Ruhepotential nähert, ist zelltypisch und durch Lernprozesse nicht veränderbar und wird durch den Parameter *Abnahmerate des EPSP* beschrieben.

Ebenso zelltypisch und unveränderlich ist der Parameter *Schwellenwert*.

Der Parameter *Signallaufzeit* fasst mehrere biologisch unterscheidbare Zeitdauern zusammen. Man kann die Signallaufzeit verstehen als Zeitspanne zwischen der

Spitze eines Aktionspotenzials in der vorgeschalteten Zelle und der Spitze eines dadurch ausgelösten Aktionspotenzials in der nachgeschalteten Zelle. Die Signallaufzeit ist also pro Verbindung, nicht pro (vorgeschalteter) Zelle zu definieren. Sie ist in der Simulation konstant, wird also durch die simulierten Lernprozesse ebenfalls nicht verändert.

Da das Aktionspotenzial ein Alles-oder-nichts-Signal ist, können die *Synapsengewichte* gleichgesetzt werden mit der Wirkung, die ein Aktionspotenzial auf das Membranpotenzial der nachgeschalteten Zelle hat. Es wird angenommen, dass die Synapsengewichte (nicht für alle Verbindungen) durch Lernvorgänge und Vorgänge des Vergessens veränderbar sind, dass es aber einen durch mehrere biologische Faktoren bedingten Maximalwert gibt, der für verschiedene Synapsentypen verschieden ist. Obwohl angenommen wird, dass eine einzelne Zelle nur *einen* Stoff als Neurotransmitter produziert, müssen die maximalen Gewichte der Synapsen, die diesen Transmitter verwenden, nicht gleich sein.

Die *Dauer der Refraktärphase* ist eine zelltypische Konstante.

Da die *Lernparameter* eine Auseinandersetzung mit der Frage der Repräsentationsform sprachlicher und nichtsprachlicher Informationen im Gehirn voraussetzen, werden sie erst in den folgenden Kapiteln an Ort und Stelle behandelt. Es gibt Lernparameter, die zellspezifisch und solche, die synapsenspezifisch sind.

Im Folgenden sollen nun einige Simulationen, der Übersichtlichkeit halber mit nur wenigen Zellen, vorgeführt werden, die vor allem den Zweck haben, zu illustrieren, wie durch das verwendete Simulationsprogramm elementare neuronale Vorgänge symbolisiert werden. In allen Beispielen ist der Zeittakt mit einer Zeitdauer von ungefähr 1ms zu veranschlagen.

Ein einfaches Beispiel für eine neuronale Funktion ist die zeitliche Summation: Nacheinander durch denselben Eingang (nicht eine notwendige Bedingung!) erregte EPSPs addieren sich und können, falls die Summe überschwellig wird, ein Aktionspotenzial auslösen. Der zeitliche Abstand der Eingangssignale ist dabei, genau so wie die Effektivität der beteiligten Synapse(n), ein wichtiger Faktor, abhängig von der Geschwindigkeit, mit der unterschwellige EPSPs in der betreffenden Zelle abfallen.

Die Abbildung 2.1.4–2 erklärt die auf dem Bildschirm erscheinende Symbolik. Es wird nur eine Auswahl der in die Berechnung eingehenden Werte dargestellt, vor allem solche, die sich im Lauf der Simulation verändern können oder aus denen man leicht Rückschlüsse ziehen kann auf Veränderungen von Werten.

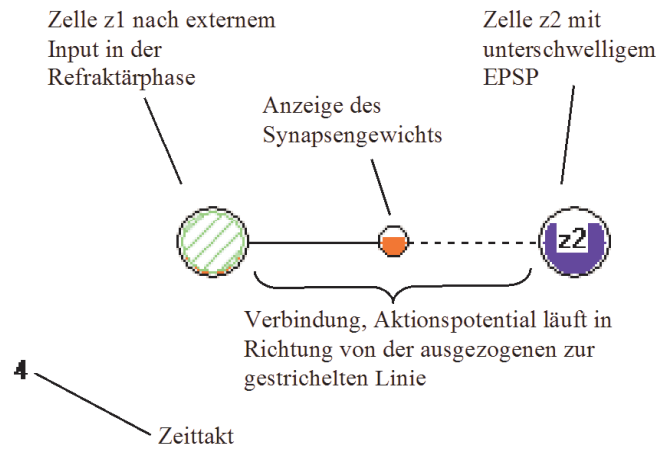


Abbildung 2.1.4–2: Zeitliche Summation, Zeldarstellung. Zellen werden durch große, Verbindungsgewichte durch kleine Kreise wiedergegeben, die Ausfüllungen stellen die jeweiligen Zustände dar. Die (grün) schraffierte Zelle ist im Augenblick in der Refraktärphase. Sie hat in Zeittakt 1 einen überschwelligen Input (blauer „Flüssigkeitspegel“) erhalten und hat in Zeittakt 2 gefeuert, das heißt ein Aktionspotential generiert (rote Ausfüllung des Zellkreises). Dieses Aktionspotential hat in der Zelle rechts ein EPSP erzeugt, kenntlich wieder an dem blauen Flüssigkeitspegel. Der kleine Kreis zeigt ein unterschwelliges Synapsengewicht an.

Simulation:

Zeitliche Summation, Zeldarstellung

Diese und die folgenden Simulationen sollten durch Betätigen und ggf. Festhalten der Leertaste gesteuert werden. Der Zustand der Zellen ändert sich nach Zeittakt 380 nicht mehr.

In vielen Fällen ist es sinnvoll, nicht nur ein filmartiges Abrollen der Vorgänge an einer Zellpopulation Zeittakt für Zeittakt zu haben, sondern eine übersichtliche Darstellung eines längeren Verlaufs, etwa in der Art, wie man Folgen von Aktionspotentialen, die durch intrazelluläre Ableitungen gewonnen werden, darstellt. Die durch das Simulationsprogramm verwendete Symbolisierung wird in Abbildung 2.1.4–3 erläutert. Im Unterschied zu der gängigen Darstellungsweise für Ableitungen ist die Zeitachse aus praktischen Gründen senkrecht gestellt. Es wird pro Zeittakt (wenn nichts anderes bestimmt ist) eine Pixelzeile geschrieben.

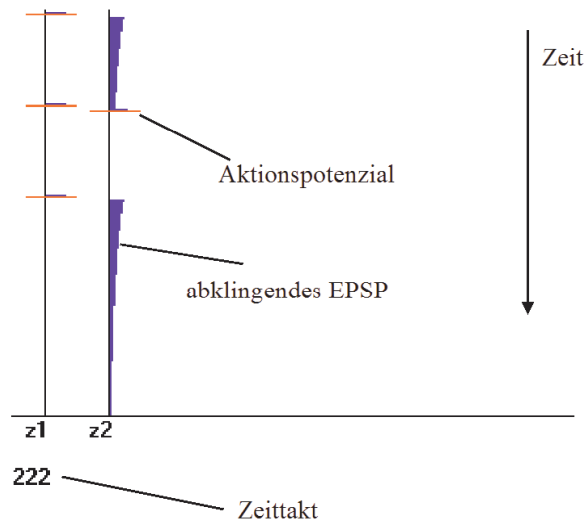


Abbildung 2.1.4-3: Zeitliche Summation, Verlaufsdarstellung. EPSPs werden durch waagrechte blaue Linien (die sich zu Flächen zusammenschließen) rechts von der senkrechten schwarzen Linie dargestellt, IPSPs entsprechend durch waagrechte Linien links. Aktionspotenziale haben weder eine auswertbare zeitliche Erstreckung noch eine auswertbare Amplitude. Sie werden deshalb einfach durch einen waagrechten Strich dargestellt, der die senkrechte Linie durchkreuzt.

Simulation:

Zeitliche Summation, Verlaufsdarstellung

Man beachte, dass der Bildschirm, solange die Leertaste nicht betätigt wird, leer bleibt.

Ein für den Kortex realistischerer Vorgang (verglichen mit der einfachen zeitlichen Summation) ist – so möchte man meinen – die Integration mehrerer, nicht synchronisierter Eingänge. Wenn verschiedene Signalintensitäten verrechnet werden sollen, so können diese Intensitäten in der einfachsten Form durch verschiedene Frequenzen von Aktionspotenzialen dargestellt werden. Das Integrationsergebnis an einem Neuron, das auf diese Weise kodierte Signalintensitäten als Input erhält, sollte wieder eine Folge von Aktionspotenzialen sein, mit einer vom Input abhängigen und prinzipiell aufgrund des Inputs vorhersagbaren Frequenz.

Wir verwenden zur Demonstration eine Anordnung von vier Zellen, die ergänzt wird durch drei auf dem Simulationsbildschirm nicht sichtbare Zellen, die die Eingabefrequenzen generieren.

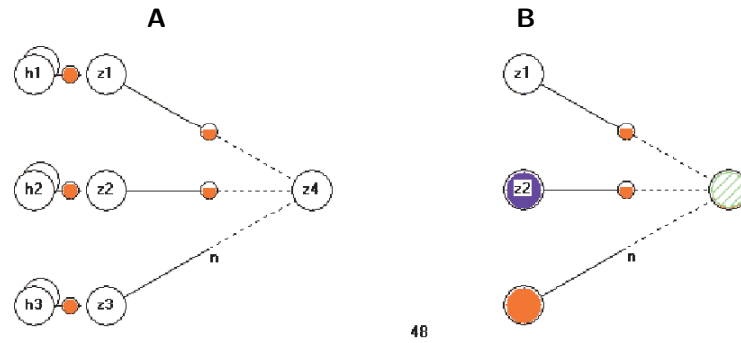
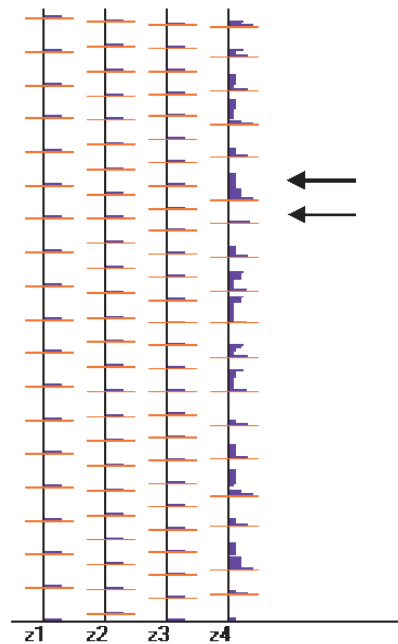


Abbildung 2.1.4-4: Integration mehrerer Eingänge, Zeldarstellung. **A.** Architektur einschließlich Zellen, die die Eingabefrequenzen durch Verbindungen unterschiedlicher Laufzeiten auf sich selbst zurück generieren (eine Form von „kreisender Erregung“). **B.** Darstellung des Simulationsbildschirms: Die Zelle $z4$ ist in der Refraktärphase (grüne Schraffur), die Zelle $z3$ feuert (rote Ausfüllung), die Zelle $z2$ hat ein überschwelliges Membranpotenzial. Die Zelle $z3$ ist eine hemmende Zelle, die Verbindung zu $z4$ ist mit einem n gekennzeichnet. (Da hemmende Verbindungen unveränderlich sind, wird ihre Effektivität nicht genauer ausgewiesen.)

Simulation:
Integration mehrerer Eingänge, Zeldarstellung

Das in Abbildung 2.1.4-5 wiedergegebene Simulationsergebnis ist in mehrfacher Hinsicht bemerkenswert. Obwohl die Impulszüge, die auf die Zelle $z4$ einwirken, völlig regelmäßig sind, ist das Ergebnis der Integration sehr unregelmäßig. Man vergleiche die Abstände der Aktionspotenziale an der Stelle der zwei Pfeile. Die Frequenz, wenn man die Aktionspotenziale über einen längeren Zeitraum zählt, entspricht ungefähr der langsamsten Eingangsfrequenz. Auch das würde man vielleicht nicht ohne weiteres erwarten, da der hemmende Eingang (bei zwar etwas größerer Frequenz) nur halb so stark ist wie jeder der beiden erregenden Eingänge.



346

Abbildung 2.1.4–5: Integration mehrerer Eingänge, Verlaufsdarstellung. Weitere Erläuterungen im Text.

Simulation:

Integration mehrerer Eingänge, Verlaufsdarstellung

Das Problem eines solchen Ergebnisses besteht vor allem in den Bedingungen, die es einer möglichen Weiterverarbeitung auferlegt. Wenn die Frequenz eine Bedeutung hat, muss offenbar eine längere Strecke des Output-Impulszugs abgewartet werden, ehe ein Weiterverarbeitungsergebnis zu irgendwelchen Konsequenzen führen kann. Es genügt nicht, den Abstand zweier aufeinanderfolgender Aktionspotenziale auszuwerten. Damit müssten die beobachtbaren und möglichen Frequenzen an kortikalen Neuronen z. B. deutlich höher sein als die dadurch gesteuerten motorischen Ereignisse. Wenn man sprachliche Phänomene betrachtet, stößt man auf die Tatsache, dass leicht 15 Laute (Phoneme) pro Sekunde produziert und im Verstehensprozess verarbeitet werden können. Man wird aber schwerlich mit einer Fre-

quenz von vielleicht 150/s an einzelnen Neuronen im Kortex rechnen dürfen. Die tatsächlich gemessenen Werte liegen deutlich darunter.

Eine Idee, die man in diesem Zusammenhang haben wird, ist, dass ja die Verrechnung weniger Eingänge vielleicht nicht typisch ist für die Funktionsweise des Kortex und dass die Unregelmäßigkeit des Outputs möglicherweise verschwindet, wenn man die Zahl der integrierten Eingänge erhöht und die Synapsenstärken entsprechend reduziert. Diese Idee ist in der folgenden Simulation realisiert. Die Zahl der Eingänge in die integrierende Zelle (Identifikation zx) ist auf zehn erhöht. Fünf der Eingänge sind hemmend. Die Synapsenstärken sind gegenüber dem vorigen Experiment halbiert. Die verwendeten Frequenzen liegen, wenn man die Dauer eines Zeittakts mit 1ms veranschlagt, zwischen 50/s und 100/s.

Wie die Abbildung 2.1.4–6 zeigt, ist der Output aber nicht deutlich regelmäßiger. Die Erhöhung der Zahl der Eingänge so weit, dass man sich noch eine einigermaßen verlässliche Verrechnung vorstellen kann, scheint das Problem nicht zu beseitigen.

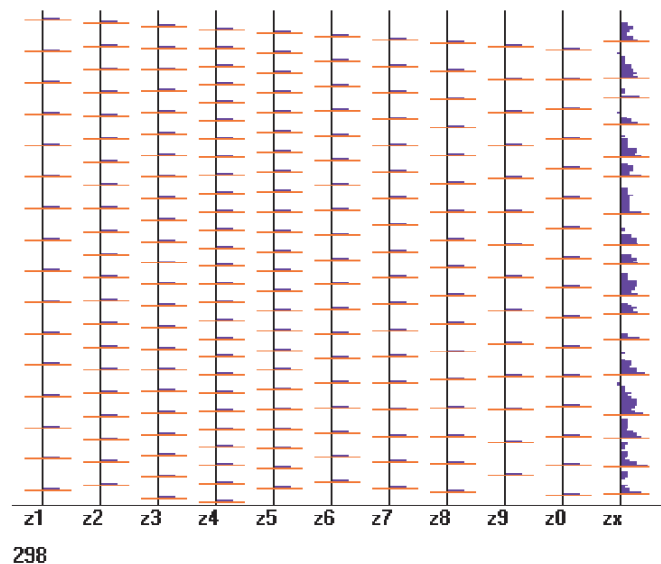


Abbildung 2.1.4–6: Beispiel einer Integration mit höherer Zahl von Eingängen. Weitere Erläuterungen im Text.

Simulation:

Erweiterte Integration, Verlaufsdarstellung

Wenn man zusätzlich die Zelldarstellung dieser Simulation verwendet, sieht man, dass sehr verschiedene „Koalitionen“ feuender Input-Zellen zu verschiedenen Zeitpunkten „zufällig“ zustande kommen, was zu ebenso „zufälligen“ Wirkungen führen muss.

Simulation:

Erweiterte Integration, Zelldarstellung

Das Fazit aus dieser Simulationsserie muss sein, dass die Auswertung von frequenzkodierter Information im Nervensystem nur eine beschränkte Bedeutung haben kann. Sie ist an das Vorkommen höherer Frequenzen gebunden und man kann sich schwer vorstellen, dass mehrere aufeinanderfolgende Verarbeitungsschritte auf Frequenzkodierung basieren.

2.2 Die Repräsentationsproblematik

2.2.1 Allgemeine Bedingungen für mentale Repräsentationen

Wenn man verschiedene Standpunkte zur Repräsentationsproblematik vorstellen und diskutieren möchte, ist es vorteilhaft, sich im Vorfeld Perspektiven zu überlegen, die man einsetzen kann, um sich nicht in der Vielfalt der Möglichkeiten zu verlieren.

Es muss zunächst festgehalten werden, dass die Behandlung des Themas in unserem Zusammenhang auf den Bereich kortikaler Repräsentationen beschränkt ist oder auf Bereiche, die sich darauf beziehen lassen. Damit sind einerseits Ideen, die für datenverarbeitende Maschinen entwickelt worden sind, eingeschlossen und andererseits der gesamte sozialwissenschaftliche Bereich ausgegrenzt. Es wird bewusst, um die Vorstellung nicht in falsche Bahnen zu lenken, nicht von Kodierung gesprochen.

Man geht am besten von einem anschaulichen Beispiel aus. Als solches soll hier die Funktion eines elektromagnetischen Relais herangezogen werden. Eine einfache Realisation eines solchen Relais sieht so aus wie in Abbildung 2.2.1-1 schematisch dargestellt.

Das Funktionsprinzip besteht darin, dass ein Steuerstrom einen Elektromagneten erregt, der damit einen oder mehrere Kontakte betätigt. Es kann z. B. so sein, dass der Steuerstrom schwach ist und damit ein viel stärkerer Arbeitsstrom geschaltet werden soll. Die Schalterstellung ist also abhängig von dem Strom in dem steuernden Stromkreis. In erregtem Zustand wird bei dem Relais der Abbildung 2.2.1-1 der Arbeitskreis geschlossen. Wird der Strom im Steuerkreis abgeschaltet, fällt das Relais ab, das heißt, es geht in einen Ruhezustand über, in Abbildung 2.2.1-1 wird dann der Arbeitskreis geöffnet.

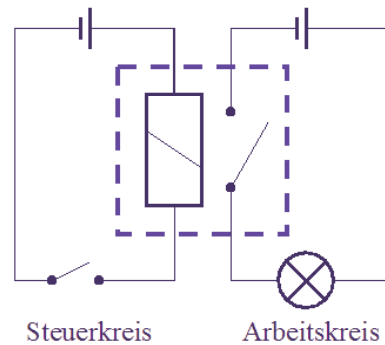


Abbildung 2.2.1-1: Schaltbild eines elektromagnetischen Relais, einbezogen in eine beispielhafte Anwendung. Innerhalb der gestrichelten Linie ist das Relais selbst dargestellt, so wie es als Bauelement im Handel ist. Das Rechteck mit dem schrägen Strich symbolisiert einen Elektromagneten, der, je nachdem ob der Steuerkreis geschlossen ist oder nicht, den danebenliegenden Schalter betätigt.

Man kann jetzt sagen, dass der aktuelle Zustand des Schalters den aktuellen Zustand des Steuerkreises repräsentiert. Das betrifft zunächst das Relais im Betrieb. Das Fließen eines Stroms im Arbeitskreis zeigt aktuell das Fließen eines Stroms im Steuerkreis an. Ein Zustand des Steuerkreises ist in diesem Fall „aktuell“ repräsentiert.

Man kann aber auch losgelöst davon, dass das Relais in Betrieb ist, sagen, dass ein geschlossener Arbeitsstromkreis einen geschlossenen Steuerstromkreis repräsentiert usw. Die Zuordnung von Schalterstellungen und Spulenerregung ist prinzipiell variabel. Es gibt Relais, die als Schließer (wie das der Abbildung 2.2.1-1), als Öffner, als Wechsler usw. arbeiten. Man kann also unterschiedliche Repräsentationen für das Phänomen „Spule erregt“ haben. Nur im Fall einer solchen prinzipiellen Variabilität wird man von Repräsentation sprechen wollen. Die Art der Repräsentation ist auch bei einem Relais außer Betrieb festgelegt. Man kann davon sprechen, dass ein bestimmtes Relais bestimmte Repräsentationen für verschiedene Zustände des Steuerkreises bereithält. Die möglichen Zustände sind in diesem Sinne „potenziell“ repräsentiert.

In einem noch grundsätzlicheren Sinn kann gesagt werden, dass die Menge der möglichen Repräsentationen durch die zur Verfügung stehende „Hardware“ beschränkt wird. Man kann mit einem elektromagnetischen Relais ohne weitere Zutaten keinen in Zeiträumen von Sekunden kontinuierlich steigenden oder fallenden Arbeitsstrom hervorbringen.

Schließlich ist es auch interessant, zu beachten, dass die Aktualisierung einer potenziellen Repräsentation einen Mechanismus braucht, der im Fall des Relais in der Magnetwirkung usw. besteht.

Es sind also vier Gegenstände zu unterscheiden:

1. Ein steuernder oder jedenfalls primärer Zustand/Vorgang;
2. eine apparative Grundlage und apparative Funktionen;
3. potenzielle Repräsentation(en);
4. aktuelle Repräsentation(en).

Zurückliegend ist ein technisches Beispiel betrachtet worden, das vor allem Aussagen bringen sollte über die Natur von Repräsentationen schlechthin. Das technische Beispiel ist nicht als Modell für mentale Repräsentationen aufzufassen, sondern als einfacher Repräsentationsfall, der einige Überlegungen grundsätzlicher Art erleichtert. Wie sieht es, in Anwendung dieses Beispiels, mit den speziellen Bedingungen für mentale Repräsentationen aus?

Es ist für mentale Repräsentationen nicht zu verlangen, dass sie einer Vereinbarung unterliegen, etwa so, wie man sagt, dass die Zuordnung von Ausdruck und Inhalt sprachlicher Zeichen einer Konvention entspricht. Mentale Repräsentationen repräsentieren sozusagen aus apparativen Gründen (genauso wie die Komponenten des Relais), nicht weil es eine äußere Vereinbarung darüber gibt. Wir können die Relais-Komponenten interpretieren, als ob es eine Vereinbarung gäbe, sie funktionieren aber ohne unsere Interpretation. Mentale Repräsentationen sind allerdings veränderbar (die Zuordnung primärer Zustand – potenzielle Repräsentation ist veränderbar), und zwar in diesem Fall, anders als bei kommerziellen Relaisstrukturen, durch Lernprozesse. Veränderung bedeutet aber nicht Veränderung einer Vereinbarung.

Mentale Repräsentationen müssen einen ausreichenden Grad von Eindeutigkeit haben. Im äußersten negativen Fall könnten mentale Repräsentationen für beliebige primäre Zustände gleich sein. Das technische Beispiel des Relais würde in diesem Fall den Arbeitsstromkreis nicht schließen, gleichgültig, ob ein Strom im Steuerkreis fließt oder nicht.

Mentale Repräsentationen müssen auf einen vorhandenen Apparat passen. Mit einem elektromagnetischen Relais lassen sich, wie oben festgestellt, nicht beliebige Repräsentation realisieren. Die Repräsentationen müssen außerdem aktualisierbar sein und müssen weitere Konsequenzen haben können

(ebenfalls eine apparative Voraussetzung). Repräsentationen sind grundsätzlich hardwareabhängig.

Potenzielle Repräsentation bilden eine Form von Gedächtnis. Potenzielle Repräsentationen können prinzipiell auch einfach im Apparat stecken, der aktualisierte Repräsentationen hervorbringt, und sind dann unter Umständen nicht vereinzelbar.

Mentale Repräsentationen sind nicht einsehbar, weder für den Besitzer noch für einen Außenstehenden. Es ist nicht notwendig, zu verlangen, dass sie von Person zu Person auch nur annähernd gleich sein sollen (das gilt mindestens in einiger Entfernung von der Sinnesperipherie bzw. der motorischen Peripherie).

Der hier verwendete Repräsentationsbegriff ist elementarer, als der von Newell in Newell (1990: 59) konstruierte:

„This is the essence of representation – to be able to go from something to something else by a different path when the originals are not available. We can cast this as a general law:

The representation law:

$$\text{decode}[\text{encode}(T(\text{encode}(X)))] = T(X)$$

where X is the original external situation and T is the external transformation“

Newells Formulierung erinnert an das Prinzip der Modellbildung, wie es in Teil 1 behandelt worden ist. Mentale Repräsentation kann aber nicht beschränkt werden auf eine Ersatzfunktion. Wenn man in diesem Zusammenhang an die Manipulation von mentalen Modellen denkt, muss darauf hingewiesen werden, dass dabei Vorstellungsprozesse beteiligt sind, die zwar auf der Basis von Repräsentationen ablaufen, aber nicht einfach identisch sind mit der Aktivierung von Repräsentationen. Man beachte auch, dass es im Falle mentaler Repräsentationen schwierig ist, davon zu sprechen, dass jemand da ist, der diese Repräsentationen zu einem bestimmten Zweck verwendet (Newell, 1990: 60). Wenn von Kodierung und Dekodierung die Rede ist, wird die Idee eines Codes als Basis des Übersetzungsprozesses zwar nicht notwendig impliziert, aber doch begünstigt, und die Grenze oder überhaupt die grundsätzliche Differenz zwischen Innen und Außen wird überbetont: Das Innen ist weniger Realität als das Außen.

2.2.2 Symbolverarbeitende Modelle

Wenn man mentale Repräsentationen behandeln möchte, ist eine wichtige Frage die, ob man von einer symbolverarbeitenden Architektur ausgehen darf, ob also die in sog. „symbolverarbeitenden Modellen“ verwendeten Strukturen modellhaft real sind mit Bezug auf mentale Strukturen. Von symbolverarbeitenden Modellen wird vor allem dann gesprochen, wenn man den Gegensatz zu nicht-symbolverarbeitenden, nämlich konnektionistischen Modellen betonen will. Aus linguistischer Sicht ist mit diesem Gegensatz z. B. die Frage verbunden, ob es mentale Repräsentationen von grammatischen Regeln gibt, oder ob die Formulierung von Regeln durch den Linguisten nur eine beobachtbare Regelmäßigkeit sprachlicher Ausdrücke wieder spiegelt.

Symbolverarbeitende Modelle menschlicher Intelligenz sind aus Erfahrungen mit computergestützter künstlicher Intelligenz heraus entwickelt worden und beruhen in vielfacher Weise auf der Annahme, dass strukturelle Eigenschaften von Computern allgemeine Anforderungen an intelligente Systeme spiegeln.

Dieser Hintergrund wird schon bei der Fassung grundlegender Begriffe deutlich. In dem folgenden zusammenfassenden Zitat aus Newell & Simon (1981: 64) wird von „patterns“ gesprochen, die aus Symbolen bestehen, und es wird ein Zusammenhang beschrieben zwischen diesen patterns und Prozessen, die durch sie ausgelöst werden und die diese oder andere symbolische Strukturen manipulieren können.

„Symbol systems are collections of patterns and processes, the latter being capable of producing, destroying, and modifying the former. The most important properties of patterns is that they can designate objects, processes, or other patterns, and that when they designate processes, they can be interpreted. Interpretation means carrying out the designated process. The two most significant classes of symbol systems with which we are acquainted are human beings and computers.“

Für Computer, die eine Von-Neumann-Architektur haben, ist die Unterscheidung von Speicher- und Verarbeitungsstrukturen konstitutiv. Auch die in dem Zitat spürbare übergeordnete Funktion der Verarbeitungsprozesse und der dahinterstehenden Apparaturen ist typisch.

Der Symbolbegriff wird in Newell (1990) noch etwas genauer charakterisiert. Symbole (genauer: Tokens von Symbolen), wie sie Newell auffasst, haben die Funktion, den Zugriff auf separat gespeicherte Daten („distal da-

ta“) zu ermöglichen. Zugriff schließt „access“ und „retrieval“ ein. „Access“ ist wohl im wesentlichen als ein Adressierungsvorgang zu sehen, der von einer zentralen Verarbeitungsinstanz ausgeht, „retrieval“ ist die Übermittlung der gespeicherten Daten an diese zentrale Verarbeitungsinstanz. Es ist in unserem Zusammenhang wichtig, sich klar zu machen, dass damit eine Unterscheidung impliziert ist zwischen einem Speicherort und einem gespeicherten Inhalt. Eine Adresse identifiziert den Speicherort, an dem der Inhalt lokalisiert ist. Der Inhalt ist, metaphorisch gesprochen (!), gegenüber dem Speicherort beweglich, kann dem Speicherort entnommen und an einen anderen Speicherort bzw. in eine Verarbeitungszentrale transportiert werden. Es gilt die „Topfmetapher“, veranschaulicht in Abbildung 2.2.2-1:

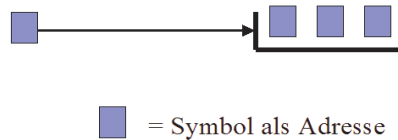


Abbildung 2.2.2-1: Trennung von Speicherplatz und Inhalt, Symbole als Adressen.

Dabei ist der „Topf“ nicht als potenzielle Repräsentation im Sinn des vorigen Abschnitts zu verstehen, denn es gibt keinen spezifischen Zusammenhang mit dem gespeicherten Inhalt. Der Topf ist inhaltsneutral (wenn man davon absieht, dass die Menge der überhaupt darstellbaren Inhalte durch die Größe und möglicherweise auch andere Eigenschaften des Topfes bestimmt wird).

Insofern als die Symbole einen Adressierungsvorgang ermöglichen, sind sie Adressen. Symbole sind aber nicht grundsätzlich als Adressen zu verstehen. Ihre Geltung würde sich sonst auf den Bereich innerhalb eines Systems beschränken, ein Außenkontakt bzw. eine Außenwirkung des Systems wäre unmöglich (vgl. Newell, 1990: 79). Die Konsequenzen werden klar, wenn man als einfaches (und vereinfachtes) Beispiel eine elektronische Schreibmaschine betrachtet. Sie besteht aus einer Tastatur, einer Verarbeitungselektronik und einem Druckwerk, z. B. einem Typenradmechanismus. Die Elektronik adressiert die einzelnen Tasten der Tastatur in einem sich wiederholenden Zyklus. So weit reicht die reine Adressierung. Was aber interessiert und was weiterverarbeitet wird, ist die Tatsache, dass eine bestimmte Taste gedrückt ist. Der Tastendruck ist ein Inhalt, der intern dargestellt wird und der nicht den Status einer Adresse hat. Ebenso hat der numerische Wert, der den mit einer Adresse versehenen Typenradmechanismus steuert, nicht den Status einer Adresse.

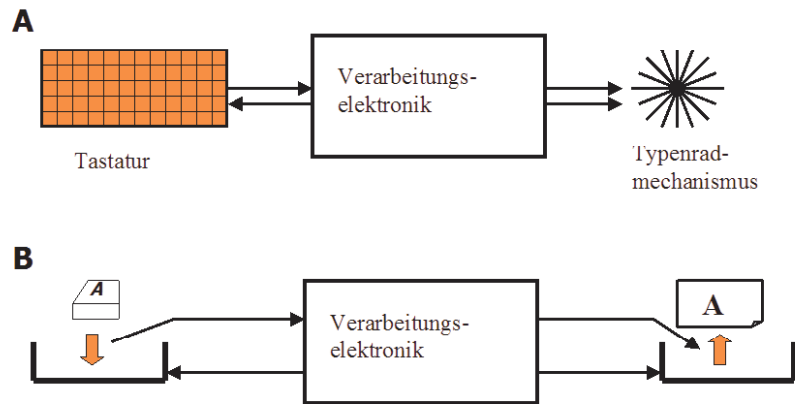


Abbildung 2.2.2-2: Adressen und Daten am Beispiel einer elektronischen Schreibmaschine. **A.** Schematische Übersicht. **B.** Ein Tastendruck erzeugt ein Datenelement (mit der Bedeutung „gedrückt“), das von der Adresse der Taste verschieden ist. Der Typenradmechanismus wird von der zentralen Verarbeitungselektronik adressiert und Daten zur Steuerung des Typenrads werden übergeben.

Es ist also für Symbole nicht notwendig, dass sie Adressen sind, sie können aber Adressen sein (zusätzlich gilt, dass eine Adresse Datenstatus erhalten kann). Charakteristisch für symbolverarbeitende Systeme ist die Unterscheidung von Daten und Adressen. Gleichbedeutend, von der Hardware her gesehen, ist die Unterscheidung von Speicherplatz und Inhalt (Topfmetapher, siehe oben). Man beachte, dass die Unterscheidung von Daten und Adressen nicht nur eine terminologische Spitzfindigkeit ist, sondern Daten werden in der gängigen Computer-Hardware anders behandelt als Adressen.

Repräsentationen in symbolverarbeitenden Systemen bestehen aus Elementen, die Paare aus Adressen und Daten sind. Abstrakt betrachtet oder jedenfalls je nach verwendeter Hardware können sowohl Daten als auch Adressen unabhängig voneinander geändert werden. Aus der prinzipiellen Koppelung der beiden Bestandteile ergeben sich einige der wesentlichsten Verarbeitungscharakteristika, die hier stichwortartig wiedergegeben werden sollen:

Wenn die Adressen geordnet sind, kann diese Ordnung den Daten sozusagen aufgeprägt werden. Es können geordnete Speicher (z. B. die für rekursive Vorgänge wichtigen Stapelspeicher, siehe unten 2.2.5) realisiert werden. Zeitliche Folgen (z. B. auch Befehlsfolgen) können auf Folgen im Adressraum abgebildet werden. Viele Lernvorgänge können als einfaches Einschreiben von Inhalten auf Speicherpositionen verstanden werden. Der Zugriff kann

dann im ungünstigsten Fall ein serielles Durchsuchen der Daten des Adressraums bedeuten.

Die Verwendung von Regeln setzt voraus, dass Variablen definiert werden können, was in einem adressierten System eine selbstverständliche Funktion ist.

Daten können von einem Speicherort auf einen anderen übertragen werden, ohne dass sie sich dabei ändern.

Es können problemlos inhaltliche Hierarchien gebildet werden. Das soll hier heißen: Adressierbare Komponenten können komplexe Einheiten bilden, die als Einheiten ihrerseits adressierbar sind.

Die Unterscheidung von aktuellen und potenziellen Repräsentationen ist wesentlicher Bestandteil. Aktuell sind adressierte Inhalte. Inhalte verschwinden nicht (notwendig), wenn sie nicht adressiert sind. Wie oben angedeutet, ist es nicht der Speicherplatz, der als potenzielle Repräsentation verstanden werden kann.

2.2.3 Konnektionismus

Es ist, wie bei den symbolverarbeitenden Modellen auch, ein Blick auf die zugrunde liegenden „Hardware“-Annahmen erforderlich.

Konnektionistische Modelle arbeiten mit Netzen von Einheiten („units“), die natürlichen Neuronen nachempfunden sind (daher „künstliche neuronale Netze“). Eine einzelne Einheit bzw. ein einzelnes „künstliches Neuron“ hat, wie das natürliche Vorbild, mehr oder weniger viele Eingangsstrukturen (die den Synapsen natürlicher Neuronen entsprechen) und nur *einen* ggf. verzweigten Ausgang (dem Axon eines natürlichen Neurons entsprechend). Die Eingänge haben unterschiedliche Effektivitäten, die durch Lernprozesse verändert werden können.

Hier endet nun allerdings die Analogie, solange man von der jüngeren Generation der „gepulsten neuronalen Netze“ („pulsed neural networks“, vgl. Maass & Bishop, 1999) absieht. Es sind vor allem die prozesshaften Annahmen, also in der Simulation diejenigen Funktionen, die das Verhalten der Neuronen beschreiben, die nur noch sehr vage die biologische Realität spiegeln. Diese prozesshaften Annahmen sind in Abbildung 2.2.3–1 dargestellt, angesichts der vielen Varianten, die es hier in der Literatur gibt, muss hinzugefügt werden: in einer Variante, die häufig in sprachverarbeitenden Modellen verwendet worden ist.

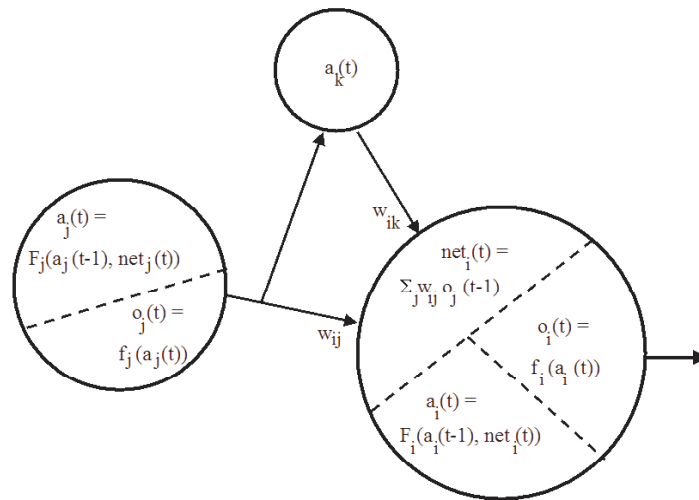


Abbildung 2.2.3–1: Schema konnektionistischer Einheiten mit Spezifizierung von Funktionen, die zur Berechnung einzelner, das Verhalten bestimmender Werte eingesetzt werden. Weitere Erläuterungen im Text. Vgl. Rumelhart, Hinton & McClelland (1986: 47) und Kochendörfer (1997: 50).

Die einzelne Zelle gibt aufgrund ihrer Aktivierung, und solange diese Aktivierung andauert (!), einen durch eine Outputfunktion bestimmten Output ab. Der Output kann in seiner Höhe – je nach Outputfunktion – den Grad der Zellaktivierung spiegeln. Die Aktivierung der Zelle wird aus der Summe der gewichteten Eingangsaktivierungen, also der Outputs der vorgeschalteten Zellen, abgeleitet. Die Gewichtung der Eingänge ist der durch Lernvorgänge veränderliche Parameter.

Konnektionistische Modelle sind „massiv parallelverarbeitende“ Modelle, das heißt, die Vorgänge können überall in einem konnektionistischen Netz gleichzeitig geschehen. Sofern die Netze auf handelsüblichen, nicht parallel, sondern seriell arbeitenden Computern simuliert werden, muss die Zeit getaktet werden. Innerhalb eines Zeittakts gelten alle auf dem Computer seriell ablaufenden Vorgänge als gleichzeitig. Im Schema der Abbildung 2.2.3–1 werden einzelne Berechnungswerte deshalb dem augenblicklichen Zeittakt t bzw. dem vorangegangenen Zeittakt $t - 1$ zugeordnet.

In der folgenden Tabelle sind einige Unterschiede zwischen dem Verhalten künstlicher und dem biologischer Neuronen zusammengestellt (wieder unter Ausklammerung der gepulsten neuronalen Netze):

Künstliches Neuron	Biologisches Neuron
Der Output ist andauernd und kann gradiert sein.	Feuern einer Zelle heißt Abgabe eines einzelnen kurzzeitigen Alles-oder-nichts-Impulses (fortgeleitetes Aktionspotenzial).
Die Aktivaton bleibt erhalten, wenn die Einheit einen Output abgibt.	Exzitatorisches postsynaptisches Potenzial (EPSP) bzw. inhibitorisches postsynaptisches Potenzial (IPSP) müssen neu aufgebaut werden, wenn die Zelle gefeuert hat.
???	Refraktärphase
Die Annahme einer Signallaufzeit ist im Allgemeinen nicht sinnvoll.	Die Signallaufzeit (Zeitspanne zwischen dem Feuern einer vorgeschalteten und dem Feuern einer nachgeschalteten Zelle) ist nicht vernachlässigbar.
Die Aktivaton einer nicht mehr von außen erregten Zelle, einschließlich des Outputs, sinkt exponentiell ab.	Eine Zelle feuert nur, wenn von außen erregt, spontan allenfalls sporadisch; das EPSP sinkt exponentiell ab, nicht die Feuerfrequenz.
Eine kontinuierliche Veränderung der Aktivaton und des Outputs ist möglich.	Es ist keine kontinuierliche Veränderung der Feuerfrequenz möglich (vgl. dazu Kochendörfer, 1997).

Nach dieser Skizze der „apparativen“ Elementareinheiten kann jetzt gefragt werden, wie Inhalte in konnektionistischen Modellen, die solche Einheiten verwenden, repräsentiert sind. Es sind prinzipiell zwei Formen zu unterscheiden, je nachdem, ob ein einzelnes Neuron als Träger eines Inhalts gesehen wird, oder ob Gruppen von Neuronen (eine „Schicht“) für diese Funktion verwendet wird. Ein einfaches, prototypisches Beispiel für die letztere Möglichkeit findet sich in Plunkett & Elman (1997), veranschaulicht in Abbildung 2.2.3-2:

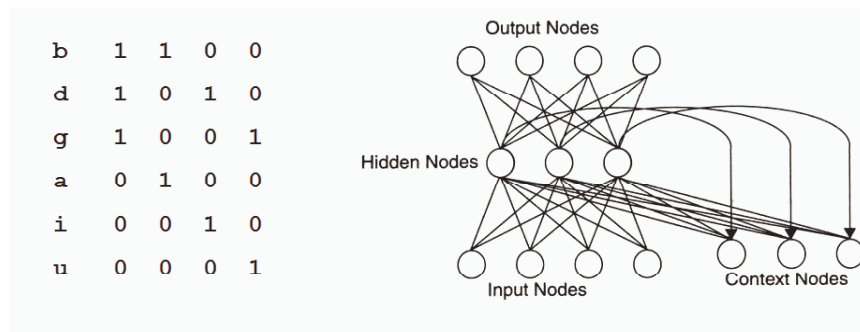


Abbildung 2.2.3–2: Beispiel für verteilte Repräsentationen. Alle Verbindungen in der Architektur rechts sind unidirektional, sie verlaufen, soweit nicht mit Pfeilspitze versehen, von unten nach oben. Weitere Erläuterungen im Text.

Das Netz modelliert Buchstabenfolgen, die aus den drei in beliebiger Reihenfolge wiederholten Ketten *ba*, *dii*, und *guu* bestehen. Eine solche Folge könnte z. B. *guugwudiidiibaguu* sein. Die erwartete Funktion ist auf die Eingabe eines Buchstabens hin jeweils die Voraussage des folgenden Buchstabens. Die Abbildung 2.2.3–2 zeigt links die per Definition festgelegten Repräsentationen für die verwendeten einzelnen Buchstaben. Sie bestehen aus Folgen von Aktivationswerten, die sich auf die vier Neuronen der Inputschicht und der Outputschicht in der rechts dargestellten Architektur beziehen. Die 1 bedeutet, dass das betreffende Neuron aktiviert ist. Das Wissen über die richtige Buchstabenfolge ist in den Verbindungsgewichten zwischen „input nodes“, „hidden nodes“ und „output nodes“ enthalten. Die Verbindungsgewichte werden über einen Lernprozess bestimmt.

Die in dem Beispiel verwendete Repräsentationsart wird als „verteilt“ bezeichnet. Da die Bezeichnung nicht immer präzise verwendet wird, ist es wichtig, sich zwei Charakteristika klar zu machen:

- Ein Inhalt (im Beispiel ein Buchstabe) wird durch ein Aktivationsmuster repräsentiert, das eine gesamte Zellschicht umfasst. Die Interpretation der Repräsentation ist nur möglich, wenn man alle Zellen der Schicht berücksichtigt.
- Wenn das Netz nicht aktiviert ist, kann man nicht sagen, dass irgendwo irgendeiner der Inhalte (Buchstaben) repräsentiert ist. Vielleicht kann man den Gesamtkomplex nach Abschluss des Lernprozesses als Repräsentation einer Problemstellung ansehen. Von dieser Möglichkeit abgesehen, kann es dann offenbar hier nur aktive Repräsentationen im Sinn von Abschnitt 2.2–1 geben.

Viele konnektionistische Modelle verwenden heute verteilte Repräsentationen in dem definierten Sinn. Die zweite, alternative Repräsentationsform wird als „lokalistisch“ bezeichnet und ist deutlich weniger populär. Die Abbildung 2.2.3–3 gibt ein einfaches Beispiel:

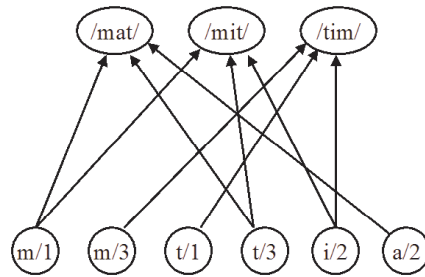


Abbildung 2.2.3–3: Beispiel für lokalistische Repräsentationen. Erläuterungen im Text.

Es sind zwei Zellschichten vorhanden, die wieder als Eingabeschicht und Ausgabeschicht verstanden werden können. In der Ausgabeschicht gibt es Zellen, die für phonologische Wörter stehen, nämlich in orthographischer Notation die Wörter *matt*, *mit* und *Timm*. In der Eingabeschicht sind die Phoneme repräsentiert, die in diesen Wörtern enthalten sind, und zwar so, dass pro Phonem und pro Position im Wort eine Zelle vorhanden ist. Das Etikett *t/3* bezeichnet ein *t* an der dritten Position in einem Wort. (In der Eingabeschicht sind der Lesbarkeit halber Schrägstriche für die Phonemgattung nicht geschrieben.) Kombinationen erregter Zellen der Eingabeschicht aktivieren entsprechende Zellen der Ausgabeschicht, und es können auf diese Weise Wortformen aus Phonemsequenzen abgeleitet werden.

Es ist bei dieser Repräsentationsform offenbar sinnlos, davon zu reden, dass eine Zellschicht, nämlich Eingabeschicht oder Ausgabeschicht, jeweils als Ganzes mit einer Bedeutung versehen ist. Eine einzelne Schicht enthält Zellen, die eher als Alternativen gelten. Wenn mehrere Zellen der Ausgabeschicht aktiviert sind, so handelt es sich um einen Fall von Mehrdeutigkeit, dasselbe gilt auch für bestimmte Kombinationen von Zellen der Eingabeschicht.

Im Unterschied zur verteilten Repräsentation kann man sagen, dass ein Konzept auch dann repräsentiert ist, wenn es nicht aktiviert ist (potenzielle Repräsentation im Sinn von Abschnitt 2.2.1).

Verteilte und lokalistische Repräsentationen ergeben unterschiedliche Gesamtarchitekturen. Dreischichtige Architekturen der in Abbildung 2.2.3–2

dargestellten Form sind typisch für eine verteilte Repräsentationsform. Das Gesamtbild der Architekturen bei lokalistischer Repräsentation ist variabler, viele Modelle arbeiten nicht mit einer festgelegten Anzahl von Schichten, und die Rede von Schichten ist oft nicht einmal sinnvoll.

In Anwendung auf mentale Repräsentationen heißt die lokalistische Repräsentationsweise spöttisch „Großmutterzellentheorie“. Die Großmutterzelle ist eine Zelle im Kortex, die dann aktiviert ist, wenn man seine Großmutter sieht oder an sie denkt. Die Großmutterzellentheorie ist oft heftig kritisiert worden. Wir kommen in Abschnitt 2.3 mehrfach wieder darauf zurück.

Konnektionistische Modelle sind nicht Implementierungen (Präzisierungen im Sinne von Teil 1) symbolverarbeitender Architekturen. Es gibt z. B. keine eigenen Strukturen zur Adressierung von Inhalten, keine Unterscheidung von (adressierbarem) Speicherplatz und einem darauf gespeicherten Inhalt. Die Architekturen und Leistungen sind insgesamt verschieden. Man kann umgekehrt selbstverständlich sagen, dass die gängigen Simulationen symbolverarbeitende Modelle sind, die als Implementierungen konnektionistischer Modelle interpretiert werden können.

2.2.4 Zellassemblies

Es sind zwei Varianten zu unterscheiden, die grundsätzlich verschiedene Funktionen annehmen und sich letztlich nur noch darin ähneln, dass die Repräsentationseinheit in dem Zusammenwirken mehrerer bis vieler Neuronen gesehen wird, und dass sich die Vertreter auf Hebb berufen, von dem der Begriff „cell assembly“ geprägt wurde.

F. Pulvermüller kann in seinen Arbeiten, die sich interessanterweise auf Phänomene der Sprachverarbeitung beziehen, als Vertreter einer noch stärker den Vorstellungen Hebbs verpflichteten Variante gelten.

Er definiert so (Pulvermüller, 1996: 24):

„Eine Cell Assembly ist eine Gruppe kortikaler Neuronen, die untereinander stark exzitatorisch verknüpft sind, wobei sich die starke Verschaltung aufgrund häufiger gemeinsamer Aktivität ausgebildet hat [...] Die starke exzitatorische Verschaltung bedingt, daß die Mitglieder einer Assembly, die Assembly-Neuronen, sich gegenseitig stark beeinflussen. Wenn ein genügend großer Teil der Neuronen aktiv ist, so wird deshalb nach kurzer Zeit die gesamte Assembly aktiviert werden. Weil dieser Aktivierungsprozeß explosionsartig

vor sich gehen dürfte, wird er auch *Zündung* genannt. Wird eine Assembly aktiviert, so kann Erregung in dem stark gekoppelten Netzwerk kreisen und so für einige Zeit erhalten bleiben. [...] Nach Hebb's neuropsychologischer Theorie sind solche Assemblies die kortikalen Repräsentanten von Gegenständen, Begriffen, Gedanken und Wörtern.“ (Vgl. auch Kochendörfer, 1999: 34.)

Die durch Lernprozesse gebildete starke exzitatorische Verschaltung, von der hier die Rede ist, kann als potenzielle Repräsentation interpretiert werden. Das „Zünden“ der Assembly stellt das aktuelle Gegenstück dar. Der Unterschied zu einer Repräsentation durch ein einzelnes Neuron besteht darin, dass erwartet wird, dass hier ein und dasselbe Neuron in verschiedenen Assemblies verwendet werden kann. Einige in unserem Zusammenhang interessierende Aspekte bleiben unklar, so z. B. die Natur des Inputs in eine Zellembly: Kann der Input prinzipiell im Output anderer, hierarchieniedriger Zellemblys bestehen oder stammt der Input immer aus lokalistischen Repräsentationen im Sinne des vorigen Abschnitts?

Die modernere Alternative der Zellembly-Theorie ist relativ ausführlich in Engel & König 1996 beschrieben (vgl. hierzu insgesamt auch Teil 1, Abschnitt 1.1.4). Zentraler Ausgangspunkt ist das „Bindungsproblem“ bei der visuellen Wahrnehmung von Objekten:

„Wahrnehmungspsychologische Untersuchungen haben gezeigt, daß die Analyse visueller Szenen mindestens zwei Verarbeitungsschritte involviert [...]. Zunächst werden in lokaler und paralleler Weise spezifische Merkmale gesehener Objekte detektiert. Diese perzeptiven Komponenten müssen dann in einem zweiten Schritt integriert und zu organisierten Einheiten gebunden werden, um die Basis für eine kohärente Repräsentation des betreffenden Objekts zu liefern.“ (Engel & König, 1996: 123)

Die vorgeschlagene Lösung des Problems besteht in der Annahme einer zeitlichen Korrelation der neuronalen Reaktion auf zusammengehörige Merkmale:

„Diesem Gedanken zufolge könnten im Kortex verteilte Neurone durch eine Synchronisation ihrer Entladungen zu Assemblies, d. h. zu kohärent aktiven Zellverbänden, zusammengeschlossen werden [...] Diese zeitlichen Korrelationen sollten dabei mit einer Präzision von wenigen Millisekunden auftreten. In einem solchen „Zeitkodierungsmodell“ würde dann die perzeptive Einheit gesehener Objekte

durch das synchrone Feuern der entsprechenden merkmals sensitiven Zellen repräsentiert.“

Hier wird deutlich, dass der Input tatsächlich lokalistisch gedacht ist. Die Synchronisation wird durch kortiko-kortikale Verbindungen gewährleistet (Engel & König, 1996: 132 und 134). Wenn die Synchronisation zur Repräsentation von Objekten dienen soll und nicht nur auf eine bildhafte Widerspiegelung (was immer das heißen mag) hinausläuft, müssen die dahinterstehenden Verbindungen durch Lernprozesse zustandekommen. Wendet man die Unterscheidung von potenzieller und aktueller Repräsentation an, bildet dieses Verbindungsgeflecht die potenzielle Basis der in dem zeitkorrelierten Feuern sichtbaren aktuellen Erscheinung.

Größere, ausreichend spezifizierte Modelle, die in unserem Zusammenhang zu beachten wären, sind unter Verwendung von Zellenassembly-Vorstellungen bisher nicht konstruiert worden.

2.2.5 Zusammenfassende Bewertung

Ellis & Humphreys (1999) formulieren in der Einleitung zu ihrer Darstellung konnektionistischer Modelle im psychologischen Bereich:

„Both Turing’s thesis and the PSS hypothesis support the view that the physical structure of the brain may have little bearing on its computational abilities. A connectionist brain may be used to realise a von Neumann virtual machine! A central tenet of connectionism amounts to a simple denial of such views in that a connectionist brain is taken to entail a connectionist mind.“ (Ellis & Humphreys, 1999:5)

Turings These besagt, dass eine sog. „Turingmaschine“ die Berechnungsvorgänge auf jedem möglichen Computer abbilden kann, die „PSS hypothesis“ („physical symbol system hypothesis“) besteht nach Newell (1980: 170) in dem Satz

„The necessary and sufficient condition for a physical system to exhibit general intelligent action is that it be a physical symbol system.“

Beide formulieren zentrale Positionen im Paradigma der Symbolverarbeitung. Die Behauptung der Hardwareunabhängigkeit ist aber nur in sehr

oberflächlicher Interpretation haltbar. Die Überlegungen in Abschnitt 2.2.2 zeigen, dass trotz gegenteiliger Versicherungen ganz bestimmte Hardwarebedingungen vorausgesetzt werden. Es ist allerdings nicht einfach, die in unserem Zusammenhang entscheidende Frage zu beantworten, wie gut die vorausgesetzte Hardware mit der „Wetware“ des Gehirns vereinbar ist.

Argumente sind in Auseinandersetzung mit dem Konnektionismus deutlich gemacht worden, z. B. bei Fodor & Pylyshyn (1988). Fodor & Pylyshyn stellen eine Reihe von Anforderungen an mentale Repräsentationen und Prozesse zusammen, die in erster Linie aus Bedürfnissen sprachlicher Intelligenz abgeleitet sind und die zugunsten einer symbolverarbeitenden Architektur ins Feld geführt werden können: „productivity“, „systematicity“, „compositionality“ und „inferential coherence“. Die letzteren drei stellen für nicht-symbolverarbeitende, aber lokalistische Architekturen kein Problem dar, wir können sie deshalb hier übergehen; die detaillierte Auseinandersetzung mit der Alternative lokalistisch vs. verteilt ist Gegenstand des Kapitels 2.3. Von besonderem, auch linguistischem Interesse ist dagegen die Produktivität. Gemeint ist das, was im Generativismus mit rekursiven Regeln zu erreichen versucht worden ist. Insbesondere geht es um „center-embedding sentences“ (Strukturen, die aus einer Anzahl von Konstituenten *a* gefolgt von einer gleichen Anzahl ggf. in weiteren Eigenschaften angeglicherer Konstituenten *b* bestehen), während andere Formen von Rekursivität (die manchmal als „Iterationen“ von „echten“ Rekursionen unterschieden werden, vgl. Reich, 1969) weniger problematisch sind. Ein klassisches deutschsprachiges Beispiel ist der Satz *Die, die die, die die Bäume beschädigen, anzeigen, werden belohnt* (Tesnière, 1959: 558). Solche syntaktischen Konstruktionen erfordern in Verarbeitungsprozessen einen sog. Stapelspeicher, das ist ein Speicher, bei dem nur das jeweils zuletzt eingespeicherte Element verfügbar ist („last in first out“). Damit wird eine Adressverwaltung vorausgesetzt, die mindestens für die üblichen konnektionistischen Systeme ein Problem darstellt. Es gibt zu diesem Rekursivitäts-Problem eine – auch von Fodor & Pylyshyn zitierte – Äußerung bei Rumelhart & McClelland (1986: 119):

„We have not dwelt on PDP implementations of Turing machines and recursive processing engines because we do not agree with those who would argue that such capabilities are of the essence of human computation. As anyone who has ever attempted to process sentences like “The man the boy the girl hit kissed moved” can attest, our ability to process even moderate degrees of center-embedded structure is grossly impaired relative to that of an ATN parser.“

Das bedeutet, dass im Konnektionismus schon die Beobachtungen, die Voraussetzung für die Annahme von Rekursivität sind, als nicht schlüssig abgelehnt werden. Damit wird allerdings auch die Chomsky'sche Unterscheidung von Grammatikalität und Akzeptabilität aufgegeben. Bei Reich (1969) wird der Vorschlag gemacht, den Stapelspeicher auf *eine* Speicherposition zu beschränken, wodurch das Problem auch für netzförmige Architekturen verschwindet. Insgesamt ist das Argument mit der Beschränktheit der Rekursion zwar interessant, aber nicht unbedingt hilfreich für die Bewertung symbolverarbeitender Architekturen, denn es wird ja nur gezeigt, dass Symbolverarbeitung nicht unbedingt erforderlich ist, nicht, dass man sie ablehnen muss. Im Gegenteil: Offenbar sind konnektionistische Modelle in einer Position, in der einem Gegenargument nur schwer beizukommen ist.

Stapelspeicher größerer Tiefe stellen kein Problem dar, wenn man, wie das bei symbolverarbeitenden Architekturen der Fall ist, mit einer Unterscheidung von Adressen und Daten arbeitet und damit eine Möglichkeit gewinnt, Daten formal zu ordnen und nach einer beliebigen Strategie auf Daten zuzugreifen. Das muss als ein nicht zu unterschätzender Vorteil für die Annahme symbolverarbeitender Architekturen erscheinen. Die hauptsächliche Motivation für die Konstruktion der damit konkurrierenden konnektionistischen Systeme besteht aber, wenn man von den Erfolgen im technischen Bereich einmal absieht, in der Idee der „brain style computation“, womit die Behauptung verbunden ist, dass symbolverarbeitende Prozesse für das Gehirn nicht realistisch sind. Also muss zur Bewertung der Konkurrenz zusätzlich dieser Aspekt beachtet werden.

Um in dieser Hinsicht mehr Anschaulichkeit zu erreichen, wird in dem folgenden Simulationsexperiment der Versuch gemacht, eine Speichereinheit aus biologischen Neuronen zu konstruieren, die Architekturdetails verwendet, wie sie in Computern üblich sind: Es werden Adressen dekodiert und davon abhängig Speicherinhalte in den Speicher eingeschrieben oder dort ausgelesen. Entsprechend den Möglichkeiten einer neuronalen Struktur werden Speicherinhalte, die ohne andauernde Aktivierung erhalten bleiben sollen, in Mustern von Verbindungsgewichten gespeichert. Der Prozess des Einspeicherns ist ein Lernprozess, der die Verbindungsgewichte entsprechend einstellt.

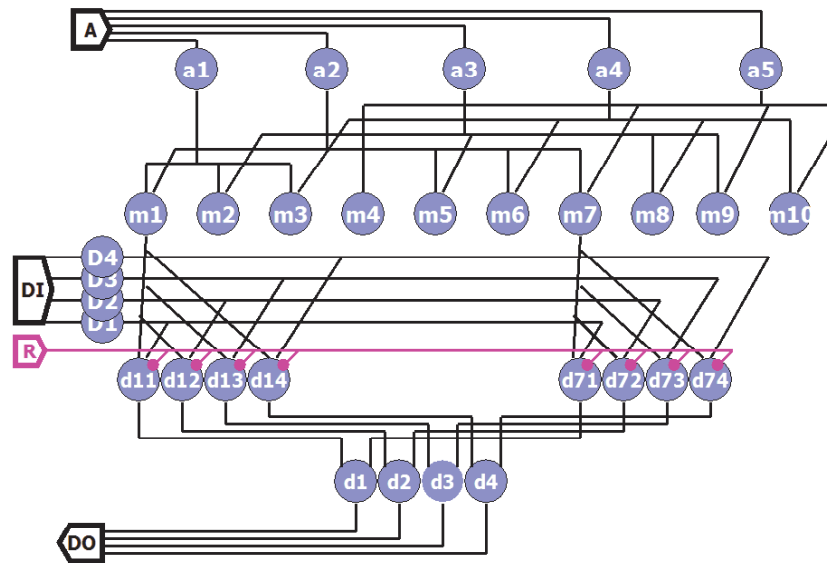


Abbildung 2.2.5-1: Skizze einer Schaltung aus biologischen Neuronen, die einen Speicher mit Adressdekodierung und Schreib- bzw. Lesevorgängen realisiert, wie er für symbolverarbeitende Prozesse vorausgesetzt wird. Das mit A bezeichnete Faserbündel gewährleistet die Repräsentation der Adressen, die als Bitmuster aufgefasst werden können. Die einzelnen Speicherpositionen werden über die mit m1 bis m10 bezeichneten Zellen angesteuert. Die gespeicherten Inhalte können wieder als Bitmuster aufgefasst werden, die vier Stellen breit sind. Der Übersichtlichkeit halber sind nur zwei der zehn Speicherplätze tatsächlich ausgeführt. Das Einspeichern der Daten geschieht über das mit DI bezeichnete Faserbündel. Da im Kortex bidirektionale Verbindungen nicht möglich sind, müssen die Datenausgänge getrennt realisiert werden (DO). Mit R ist ein hemmender Eingang bezeichnet, der beim Lesevorgang benutzt wird.

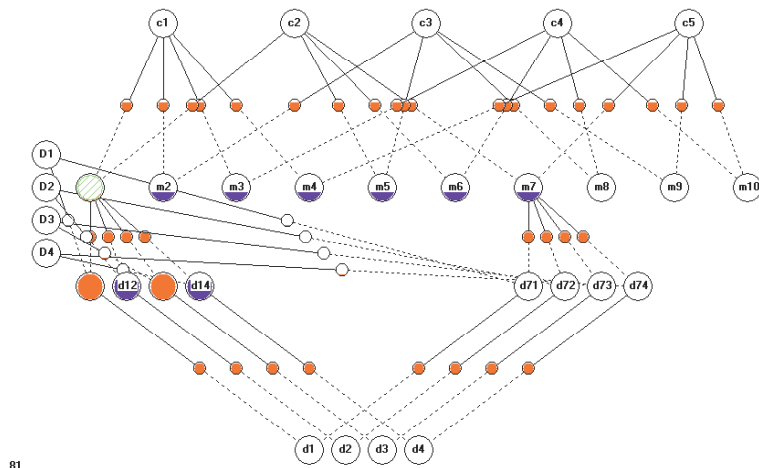
Details:

Es wird darauf verzichtet, Impulsbursts zur Repräsentation einzusetzen, alle Vorgänge basieren auf der Produktion und Verarbeitung einzelner Impulse. Die Adressierung kann dann durch Koinzidenz von Eingangsimpulsen (hier: aus genau zwei Eingängen) erzielt werden. Da das Nicht-Feuern von Zellen nicht ausgewertet werden kann, sind gegenüber technischen Realisierungen die Adressierungsmöglichkeiten beschränkt.

Schreiben: Das zu speichernde (Bit-)Muster wird an den „Dateneingängen“ angelegt, die darauffolgende Aktivierung der Adresse führt nach der Methode des

Hebb'schen Lernens (eigentlich in einer Variante, die unten in Abschnitt 2.3.2 noch näher erläutert wird) zur Verstärkung der entsprechenden Verbindungsgewichte.

Lesen: Das Ergebnis aus der Adressierung auf den eigentlichen Speicherzellen (d_{11} bis d_{14} bzw. d_{71} bis d_{74}) wird durch einen hemmenden Eingang so weit abgeschwächt, dass nur noch eine zusätzliche Erregung die Speicherzellen zum Feuern bringt. Diese zusätzliche Erregung kommt von den Dateneingängen her, die in diesem Falle alle feuern müssen.



81

Abbildung 2.2.5–2: Simulation zu der Schaltung von Abbildung 2.2.5–1. Zustand beim Lesen der in Speicherzelle m_1 eingespeicherten Information. Man kann an den kleinen Kreisen in der Mitte der Verbindungen im Bereich der Datenrepräsentation die erhöhten Synapsengewichte erkennen, in denen die Information verankert ist. Die Zelle m_1 hat gerade gefeuert, das Feuern von D_1 bis D_5 liegt schon etwas weiter zurück. Die Zellen d_{11} bis d_{14} zeigen die erwarteten Werte, die Zellen d_{71} bis d_{74} feuern nicht.

Simulation:

Adressierbare Speicherstruktur

Ergänzender Hinweis: Simulation mit Leertaste steuern!

Es ist also offenbar tatsächlich möglich, eine computerähnliche Speicherstruktur mit biologischen Neuronen zu realisieren. Es sind ganz offensichtlich auch Kriterien anwendbar, die in Abschnitt 2.2.3 auf konnektionistische Modelle bezogen worden sind: Adressen und Daten werden verteilt dargestellt, dazwischen liegt eine lokalistisch interpretierbare Schicht von Zellen, die Speicherpositionen sozusagen „etikettieren“.

Die sichere Verwendung einer verteilten Repräsentation in einem System mit natürlichen Neuronen, deren Reaktion zeitgebunden ist, würde allerdings eine Art Taktung des Gesamtsystems erfordern, ein Nachteil, der aber den Standard-Konnektionismus in gleicher Weise trifft.

Man kann natürlich, wegen der langsamen Verarbeitungsgeschwindigkeit der Neuronen, verglichen mit der von Transistoren in Computern, auch nicht annehmen, dass im Gehirn ein einziger Speicher einer einzigen Verarbeitungseinheit gegenübersteht. Um Parallelverarbeitung in dem erforderlichen Ausmaß zu ermöglichen, müssen sehr viele Speicher-Prozessor-Paare angenommen werden. Das wiederum bringt ein Problem der Kommunikation zwischen diesen Speicher-Prozessor-Paaren mit sich.

Alle diese Schwierigkeiten verblassen gegenüber der Feststellung, dass die vorausgesetzten Verbindungsmuster, z. B. die Wiederholung von gleich großen Speichereinheiten (nach der Zahl der Zellen gerechnet) oder die weitreichenden „Leitungsbündel“ fester Breite, die für die Übermittlung von Adressen und Daten erforderlich sind, im Gehirn einfach nicht nachgewiesen werden können. Bei dem aus Neuronen konstruierten Modell ist die gängige Ausrede, dass man sich eben auf einer höheren Abstraktionsebene bewege, nicht möglich. Wie auch immer man zu dem Problem der Rekursivität steht, die in symbolverarbeitenden Modellen vorgeschlagenen Lösungsvorschläge können nicht korrekt sein.

Damit ist nun nicht gesagt, dass der Konnektionismus in der in Abschnitt 2.2.3 beschriebenen Form die Architekturen der Wahl bietet. Denn es ist nicht nur so, dass es Probleme gibt mit der biologischen Korrektheit der Funktion einzelner „units“, sondern auch die Verbindungsmuster, die in typischen, vor allem verteilten Modellen verwendet werden, sind ganz offenbar unbiologisch. Man kann keine zwei oder drei vollständig verknüpften Zellschichten im Gehirn erkennen, auch nicht, wenn man zu gewissen Zugeständnissen bereit ist. Im Gehirn gibt es sehr viel mehr kurzreichweitige Verbindungen, als in verteilten konnektionistischen Modellen vorgesehen. Das Verbindungsmuster im Gehirn ist auch nicht vereinbar mit der Annahme mehrerer bis sehr vieler einzelner konnektionistischer Module. Diese Argumente gelten unter der Annahme verteilter Repräsentation. Bei lokalistischer Repräsentation sind die Architekturen beliebiger ggf. auch chaotischer, sie sind unter diesem Aspekt bessere Kandidaten.

Man wird sagen, dass es unter diesen Umständen vielleicht sinnvoller ist, gleich zu den von vornherein stärker der Biologie verpflichteten Zellensembles überzugehen. Einige der verwendeten Architekturannahmen gehen direkt auf biologische Experimente zurück, die Voraussetzungen, die für die Anatomie des Kortex gemacht werden müssen, sind nicht sofort kritisierbar.

Hier sind es Mängel in der Erklärungsleistung, die beachtet werden müssen.

Wenn die Zellenassemblies in der Version von Pulvermüller in ihrer Funktion nicht einfach identisch sein sollen mit redundanten lokalistischen Systemen, wie sie im Konnektionismus vorgeschlagen worden sind, muss eine Überlappung der einzelnen Assemblies in dem Sinn angenommen werden, dass einzelne Zellen zu verschiedenen Assemblies gehören können. Die Zuordnung der einzelnen Zellen zu den Assemblies muss durch Lernprozesse erfolgen.

Abgesehen von Problemen mit den erforderlichen Lernvorgängen, die eine Kombination von LTP (Long Term Potentiation) und LTD (Long Term Depression) vorsehen (Pulvermüller, 1999: 257), was eine geeignete, nicht störende Spontanaktivität der Zellen voraussetzt (siehe unten 2.3), kann tatsächlich eine Anordnung gefunden werden, die den Anforderungen genügt, das heißt, die eine Unterscheidung von Repräsentationen trotz Überlappung zeigt. Das wird in der in Abbildung 2.2.5–3 dargestellten Simulationsarchitektur demonstriert.

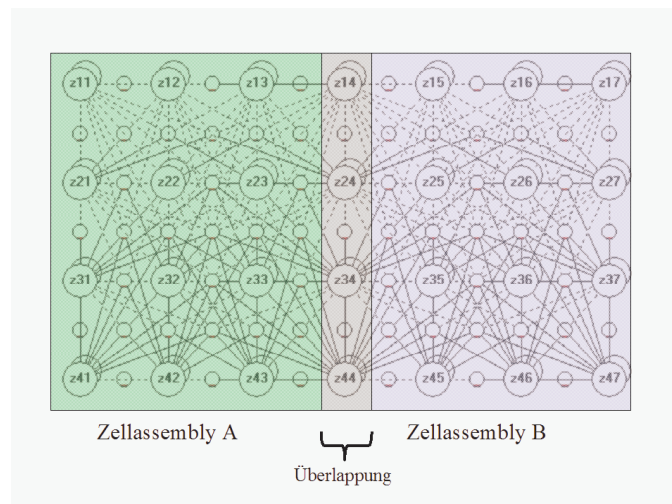


Abbildung 2.2.5–3: Anordnung zur Simulation der Erregung zweier überlappender Zellenassemblies.

Simulation:

Zellenassembly A

Zellenassembly B

Ergänzender Hinweis: Simulation mit Leertaste steuern!

Die Erregung einer der Zellenassemblies in Abbildung 2.2.5–3 mit einer Auswahl der Assembly-Neuronen führt tatsächlich zu dem erwünschten Zünden der betroffenen Assembly und nicht der überlappenden Schwester. Man muss aber beachten, um welchen Preis und mit welchen Begleitumständen das geschieht: Die Verbindungsgewichte innerhalb der Assemblies sind so eingestellt, dass die Effektivität der *gemeinsamen* Zellen, wenn sie erregt sind, nicht ausreicht, um die Assemblies zu zünden. Wie soll dieses Gleichgewicht durch Lernprozesse entstehen? Man muss auch vernachlässigen, dass nach der Zündung einer Assembly die erzielte Feuerfrequenz der einzelnen Zellen im Wesentlichen nur noch von deren Eigenschaften, nicht vom Input abhängt. Die Verbindungen zwischen den einzelnen Zellen einer Assembly müssen immer lang und langsam sein, um die im Gehirn zu beobachtenden geringen Frequenzen zu erklären. Es wird darüberhinaus nicht klar, wie das Feuer der Assembly *zuverlässig und ggf. sehr rasch* wieder abgeschaltet werden kann. Wie sollen die erforderlichen hemmenden Zellen angesteuert werden? Braitenbergs Idee des „threshold control mechanism“ entbehrt jeder direkten Grundlage in der Biologie (Braitenberg & Schüz, 1998: 197 f.; Pulvermüller, 2002).

Wenn wir zu der zweiten Variante der Assembly-Theorie übergehen, entstehen eher noch größere Probleme mit dem Verständnis der Funktion in einem größeren Zusammenhang. Es kann ja nicht nur darum gehen, einen Inhalt zu repräsentieren, sondern es muss möglich sein, diese Repräsentationen in weitergehenden Prozessen zu verwenden. Bei Ziemke & Cardoso de Oliveira (1996: 26) wird das damit angesprochene, auch schon in Engel, König & Singer (1993) zugegebene Problem (vgl. das entsprechende Zitat in Teil 1, Abschnitt 1.1.4) so formuliert:

„Selbst bei den Vorschlägen zur Lösung des binding-problems durch eine (wie auch immer geartete) Korrelation der Zeitstruktur der Aktivitäten verschiedener Neurone bleibt jedoch nach wie vor das Problem erhalten, wie diese Korrelation durch nachgeschaltete operationale Einheiten nicht nur festgestellt, sondern auch zum tatsächlichen „Verständnis“, z. B. der zu analysierenden Szene, genutzt werden kann [...].“

Aus linguistischer Sicht bedeutet das auch, dass jede Form von Hierarchiebildung, die doch zu den sichersten Befunden in allen denkbaren Teilbereichen gehört, unerklärt bleiben muss.

Kopfschüttelnd kann man dann noch den Schlusssatz bei Engel & König (1996) zur Kenntnis nehmen:

„Die am visuellen System durchgeführten Arbeiten ergeben bislang zweifellos die überzeugendsten Argumente für die Existenz zeitlicher Bindungsmechanismen, aber auch hier muß durch zukünftige Experimente auf noch direktere Weise gezeigt werden, daß die neuronale Synchronisation in kausaler Beziehung zu Wahrnehmungsprozessen steht.“ (Engel & König, 1996:148)

Das Fazit aus dem Versuch einer Bewertung der existierenden Modellbildungen zur Repräsentationsproblematik kann ungefähr wie folgt zusammengefasst werden:

Symbolverarbeitende Modelle (bzw. die dort verwendeten Repräsentationen) entfernen sich zu weit vom Gehirn, obwohl sie prinzipiell sehr detailliert und leistungsfähig sind. Repräsentationen mit Zellassemblies sind (bisher) wenig detailliert und ihre Leistungsfähigkeit lässt zu wünschen übrig, obwohl sie sehr nahe am Gehirn sind. Als wenigstens diskussionswürdig erscheinen noch am ehesten die konnektionistischen Modelle, die bezüglich beider Gesichtspunkte, Nähe zu neuronalen Strukturen und Leistungsfähigkeit, eine mittlere Position einnehmen.

Es ist festzuhalten, dass keines der Modelle den Ansprüchen, die man sinnvollerweise unter linguistischer Perspektive haben muss, genügt.

2.3 Lokalistische Repräsentation von Konzepten

2.3.1 Beobachtungen

Es ist oben in Abschnitt 2.2.3 darauf hingewiesen worden, dass die Entscheidung für eine der Repräsentationsalternativen lokalistisch vs. verteilt Konsequenzen für die Gesamtarchitektur von Systemen hat. Diese Entscheidung ist also grundlegend und muss mit der entsprechenden Sorgfalt getroffen werden. Dabei ist zu beachten, dass es unter der Voraussetzung der Präzisionsforderung (vgl. Teil 1) nicht nur darum geht, wie ein Modell konstruiert sein soll, um ein bestimmtes Verhalten zu erzeugen. Gültige modellseitige Strukturen müssen den Strukturen im Originalbereich entsprechen. Wenn solche Strukturen im Modellbereich diskutiert werden, ist gleichzeitig von entsprechenden Strukturen in der modellierten Realität die Rede, und umgekehrt. In diesem Abschnitt soll auf einige akzeptierte Erkenntnisse hingewiesen werden, die auf die Entscheidung für einen Repräsentationstyp Einfluss haben können.

Es gibt Bereiche, in denen wesentlich mehr Informationen über neuronale Strukturen bekannt sind, als das für den Kortex der Fall ist. Dazu gehören die Sinnesorgane und die äußersten Schichten der Motorik. Wenn wir uns zunächst der Motorik zuwenden, so ist bekannt, dass eine einzelne Muskelfaser bei Wirbeltieren (und damit auch beim Menschen) nur durch eine einzige Synapse erregt wird, die durch eine Axonverzweigung eines Motoneurons gebildet wird („motorische Endplatte“). Die verteilte Ansteuerung einer Muskelfaser müsste aber so erfolgen, dass mehrere Synapsen vorhanden sind, deren spezifisches gemeinsames Erregungsmuster eine Muskelfaser zur Kontraktion bringen würde. Das entspricht offenbar nicht der Realität. Verzweigungen des Axons eines Motoneurons innervieren zwar mehrere bis viele Muskelfasern, aber es ist *ein* spezialisiertes Neuron, das zusammen

mit den von ihm erregten Muskelfasern eine „motorische Einheit“ bildet. Die Erregung der Muskulatur geschieht ohne Zweifel lokalistisch und nicht verteilt.

Eine lokalistische Organisation gilt auch für die Sinnesorgane unstrittig. Niemand nimmt an, dass Mechanosensoren der Haut die Intensität des Reizes und den Reizort in verteilt-kodierter Form, das heißt über ein Erregungsmuster auf mehreren Nervenfasern übermitteln, sondern die Übermittlung der Ortsinformation geschieht dadurch, dass eine *ortsspezifische* Nervenfaser die Intensitätsinformation frequenzkodiert weitergibt. Es ist, wohlgemerkt, nicht notwendig, dass die Nervenfasern somatotop *geordnet* sind, obwohl das teilweise bis in den Kortex hinein der Fall ist. Die Kodierung ist auch ohne dieses Detail zweifelsfrei lokalistisch.

Für die Linguistik ist das Ohr als Sinnesorgan von besonderem Interesse. Die neuronale Verarbeitung des akustischen Signals beginnt in der schneckenförmig gewundenen Cochlea des Innenohrs mit den Hörsinneszellen, die nach ihrem Aussehen „Haarzellen“ heißen. Es gibt entlang der cochleären Trennwand, die vom breiten Ende des Schneckengangs bis nahe der Spitze verläuft, vier Reihen von Haarzellen. Für die Registrierung des Schallsignals ist die innere, dem Zentrum der Schnecke zugewandte Reihe von Haarzellen zuständig. Aufgrund der Form der Schnecke, der Breite und Steifigkeit der cochleären Trennwand und einer verstärkenden Mitwirkung der äußeren Haarzellen entsteht eine Frequenzaufspaltung des Schallsignals so, dass für jede Stelle der cochleären Trennwand und damit für die dort lokalisierten Haarzellen eine charakteristische Frequenz angegeben werden kann, auf die sie am besten reagieren. Haarzellen sind also auf bestimmte Frequenzen spezialisiert. Man spricht von der Tonotopie der neuronalen Struktur. Für unseren Zusammenhang gilt damit, dass die auditive Peripherie lokalistisch organisiert ist und das Ohr eine Umsetzung der akustischen Information in eine lokalistische Kodierungsform leistet. In den Handbüchern wird über die Hörbahn gesagt, dass dort die höheren Neuronen zunehmend nicht mehr auf einzelne Sinusfrequenzen, sondern auf komplexere Schallmuster reagieren, andererseits gilt das tonotope Prinzip offenbar in einzelnen Bereichen auch noch im Kortex.

Wir können anhand der vorgestellten Beispiele feststellen, dass offenbar an der Sinnesperipherie und vielleicht auch teilweise noch im Kortex die lokalistische Kodierung von Information gilt. Das bedeutet allerdings noch nicht, dass der Kortex generell mit lokalistischen Repräsentationen arbeitet. Die Zweifel betreffen vor allem die höheren Verarbeitungsebenen, z. B. auch die Sprachverarbeitung oberhalb der Lautebene oder die Lautebene auch noch einschließend. Die schematische Darstellung in Abbildung 2.3.1–1 soll das

veranschaulichen. Wenn dort das Etikett „Kortex“ gebraucht wird, bezieht sich das auf Kortexareale nicht festgelegter Ausdehnung, es soll nicht behauptet werden, dass die Frage nach der Repräsentationsform für alle Kortexbereiche als offen gelten muss. Die Sortierung der Zellsymbole in Schichten ist nur in hoher Abstraktion gültig, erleichtert es aber, in den folgenden Überlegungen einen Zusammenhang mit konnektionistischen Vorstellungen zu etablieren, der für die Argumentation wichtig wird.

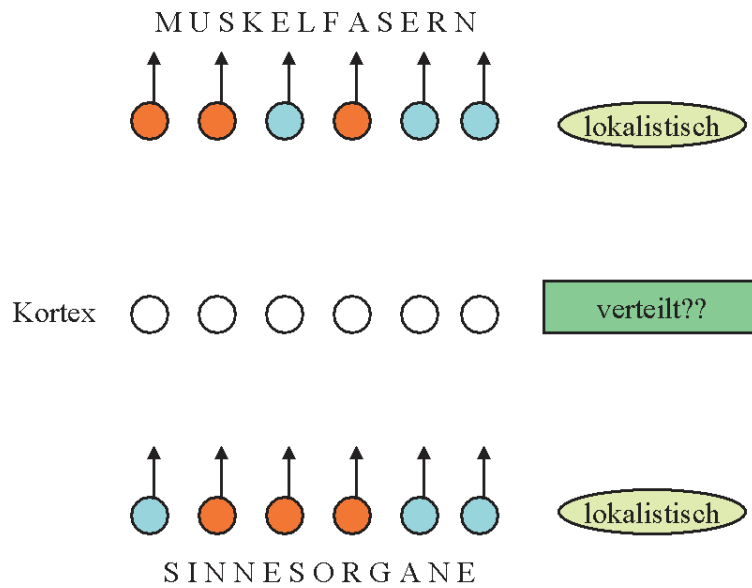


Abbildung 2.3.1-1: Veranschaulichung des Repräsentationsproblems mit Bezug auf verschiedene Bereiche des Nervensystems. Während für die Ebenen der Sinnesorgane und der Motorik das lokalistische Repräsentationsprinzip gilt, ist das für den Kortex nicht selbstverständlich.

2.3.2 Lokalistische Repräsentation, Lernprozesse und Einzelimpulskodierung

Man könnte angesichts der Verhältnisse in Abbildung 2.3.1-1 (unter der Voraussetzung, dass offenbar keine schlüssigen Argumente vorliegen, weil ja sowohl verteilte als auch lokalistische Repräsentationen in der Forschung

verteidigt werden) Ockhams Rasiermesser anwenden und sagen, dass eine Umsetzung der peripheren lokalistischen Kodierung in eine zentrale verteilte für die höhere Verarbeitung begründet werden muss, und solange das nicht geleistet werden kann, für die gesamte zentrale Verarbeitung, wie an der Peripherie auch, lokalistische Repräsentationen anzunehmen sind.

Die große Popularität der Annahme verteilter Repräsentationen lässt aber doch zögern, und so soll im Folgenden der Versuch einer zusätzlichen Argumentation unternommen werden. Dabei wird eine große Rolle spielen, dass Repräsentationen im Kortex zu einem sehr hohen Prozentsatz durch Lernprozesse zustande kommen müssen. Das sind Lernprozesse, die zunächst (aber nicht ausschließlich) über Sinnes-Inputs getrieben werden. Weil der Übergang von den internen Repräsentationen zur Motorik in dieser Hinsicht besondere Probleme aufwirft, tut man gut daran, den Weg von der Sinneswahrnehmung zur internen Repräsentation und den Weg von der internen Repräsentation zur Motorik getrennt zu behandeln. Hier interessiert hauptsächlich der Weg von der Sinneswahrnehmung zur internen Repräsentation. Der Aufbau einer internen Repräsentation muss darin bestehen, dass Verbindungen – in konnektionistischer Terminologie ausgedrückt – von der Inputschicht her so angepasst werden, dass eine sinnvolle interne Reaktion auf einen Inputstimulus entsteht.

Nicht alle Repräsentationen haben eine direkte Wirkung auf die Motorik, sodass sie über motorische Reaktionen von außen kontrolliert werden könnten. Auch sprachliches Lernen kann nicht vollständig(!) über die Produktion kontrolliert werden. Für den Weg von der Sinneswahrnehmung zur internen Repräsentation sind deshalb Formen des „supervised learning“ – wieder in konnektionistischer Terminologie ausgedrückt – ungeeignet, da biologisch unplausibel, wie auch die Konnektionisten wissen (hier für das Standardverfahren „back-propagation“ formuliert):

„... back-propagation does not seem to map onto any known brain mechanism, requiring as it does a “teacher” who computes precise error signals and passes them back down the network.“ (Ellis & Humphreys, 1999: 43)

Man beachte, dass das Problem nicht nur darin besteht, einen Output zu kontrollieren, sondern es müssen auch die entsprechenden Anpassungen der Verbindungen, die ja das Ziel des Lernvorgangs sind, irgendwie zustande gebracht werden.

Biologisch plausibel ist nur – wenn wir weiter bei den im Konnektionismus eingeführten Begrifflichkeiten bleiben – „unsupervised learning“ (das „reinforcement learning“ ist, solange man die Möglichkeit einer vollständi-

ge Kontrolle interner Repräsentationen über eine äußere(!) Reaktion ausschließt, ebenfalls unter die nicht biologisch realistischen Lernprozeduren zu rechnen).

Man kann nun beobachten, dass nicht-überwachtes Lernen offenbar, wenn man existierende Anwendungen, also z. B. konnektionistische Modellbildungen psychologischer Prozesse untersucht, immer zu lokalistischen Repräsentationen führt. Oder, anders ausgedrückt, nicht-überwachtes, also biologisch plausibles Lernen ist nur bei der Annahme lokalistischer interner Repräsentationen möglich. Bei der großen Bedeutung, die diese Feststellung hat, sollte man sich allerdings noch etwas mehr Mühe mit der Begründung machen. Es könnte ja sein, dass nicht-überwachte Lernalgorithmen für verteilte Repräsentation prinzipiell möglich sind, aber eben noch nicht gefunden wurden oder nicht verwendet werden. Die folgenden Überlegungen sollen in diesem Sinne zusätzliche Argumente bereitstellen.

Man wird sich wahrscheinlich rasch darauf einigen, dass eine besonders wichtige Teilleistung interner Repräsentationen das Wiedererkennen von Sinneswahrnehmungen ist. Wenn es nur um das Wiedererkennen geht, ist es gleichgültig, wie eine bestimmte interne Repräsentation im Detail tatsächlich aussieht, es ist nur zu verlangen, dass ein für eine Wahrnehmung stehendes Muster für diese Wahrnehmung spezifisch ist. Es darf nicht sein, dass Wahrnehmungen das gleiche Muster benutzen, wenn diese Wahrnehmungen als je verschieden identifiziert werden sollen. Wenn Repräsentationen durch Lernprozesse gebildet werden und sie nicht einfach exakte Wiederholungen der verschiedenen Inputmuster sind, sondern, wie bei verteilten Systemen selbstverständlich, eine gewisse Beliebigkeit haben, muss dafür gesorgt werden, dass eine schon aufgebaute Repräsentation nicht noch einmal für einen Lernprozess zur Verfügung steht. Das kann nur über eine Einrichtung geschehen, die, metaphorisch gesprochen, über schon vergebene Repräsentationen Buch führt und schon vergebene Repräsentationen für Lernprozesse „sperrt“. Diese Einrichtung hat aber damit die Funktion eines Tutors, es entsteht zwangsläufig eine Form überwachten Lernens. Man beachte, dass der Tutor bei bekannten Lernalgorithmen für verteilte Systeme in der Lernphase eine Liste der gewünschten Repräsentationsmuster verwendet und damit deren Distinktivität gewährleistet.

Das Fazit muss sein: Wenn nur nicht-überwachtes Lernen biologisch akzeptabel ist und nicht-überwachtes Lernen immer auf lokalistische Repräsentationen führt, muss in allen Bereichen, in denen mit Lernprozessen zu rechnen ist, eine lokalistische Repräsentationsform angenommen werden. Es ist also die Großmutterzellentheorie, die bei der Auseinandersetzung über die Repräsentationsformen den Sieg davonträgt.

Man beachte aber, dass die Argumentation bisher auf der Basis und mit Bezug auf Vorbilder im Standard-Konnektionismus geführt worden ist. Man muss nun auch noch zeigen, dass diskussionswürdige Lernverfahren auch für biologische Strukturen im engeren Sinne angegeben werden können. Wir gehen dabei zunächst noch einmal aus von einem typischen standard-konnektionistischen Lernverfahren, nämlich einer einfachen Form des „competitive learning“.

Das „competitive learning“ arbeitet mit Architekturen aus zwei Zellschichten, einer Eingabeschicht und einer Ausgabeschicht. Jede Zelle der Eingabeschicht ist unidirektional mit jeder Zelle der Ausgabeschicht verknüpft. Zusätzlich gibt es hemmende Verbindungen zwischen allen Zellen der Ausgabeschicht, sodass eine erregte Zelle jeweils alle anderen Zellen hemmt. Die Stärke der Hemmung ist abhängig von der Erregungsstärke der Zelle, von der sie ausgeht.

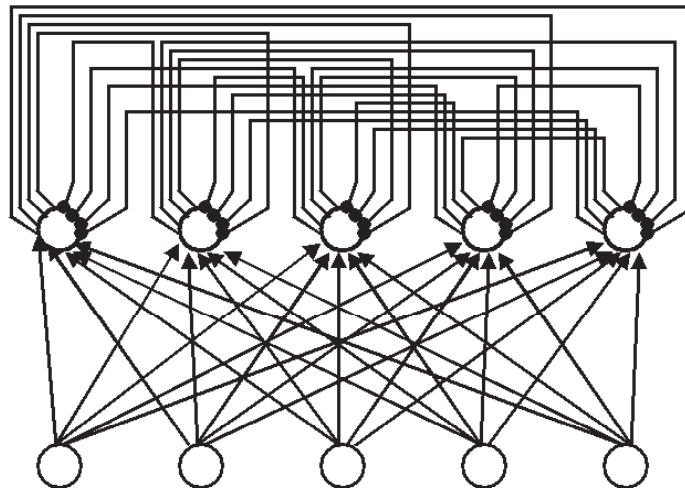


Abbildung 2.3.2-1: Architektur zum „competitive learning“ im Standard-Konnektionismus.

Das erwartete Lernergebnis besteht darin, dass auf ein einzelnes Erregungsmuster in der Eingabeschicht jeweils eine einzelne Zelle der Ausgabeschicht aktiviert wird.

Vor Eintritt in den Lernprozess haben alle Verbindungen zwischen Inputschicht und Outputschicht zufällige (positive) Werte. Der Lernvorgang be-

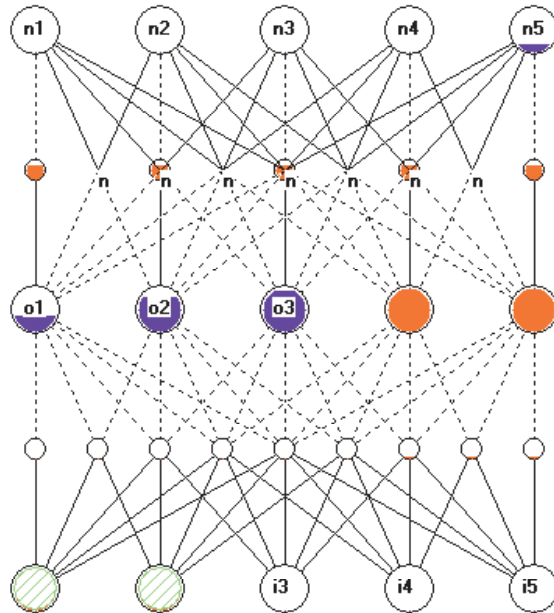
steht zunächst darin, dass ein Eingabemuster eine gewisse Zeit, das heißt über mehrere Zeittakte präsentiert wird. Während dieser Zeit findet der „Wettbewerb“ nach dem Winner-take-all-Prinzip statt, der dem Verfahren den Namen gegeben hat: Die Aktivierungen der Zellen der Ausgabeschicht verändern sich aufgrund der Wechselwirkung von Erregung und Hemmung so, dass schließlich nur noch eine einzige Zelle als Sieger des Wettbewerbs stark aktiviert ist. Wenn dieser Zustand erreicht ist, werden die Verbindungsgewichte zwischen Eingabeschicht und Ausgabeschicht so angepasst, dass die Verbindungen, die auf die nicht aktivierten Ausgabezellen führen, unverändert bleiben und die Verbindungen auf die stark aktivierte Ausgabezelle in Abhängigkeit von der Aktivierung der Eingabezellen so verändert werden, dass Verbindungen, die von aktivierten Eingabezellen ausgehen, verstärkt und die anderen abgeschwächt werden. (Letzteres führt zur Sperrung der Siegerzelle für Lernvorgänge mit Eingabemustern, die von dem aktuellen Muster verschieden sind.)

Wenn man versuchen möchte, dieses Verfahren in die biologische Realität zu übersetzen, muss einiges sofort geändert werden. Wie schon in Kochendörfer (1997: 64 ff.) festgestellt, kann ein andauernder frequenzkodierter Input keine steigende Feuerfrequenz in einem biologischen Neuron zur Folge haben. Bei Rückkoppelung ergibt sich sofort eine konstante höhere Frequenz (Lernvorgänge an den Synapsen sind in der ersten Phase des Lernprozesses nicht vorgesehen). Also kann die Aktivierung in einer potenziellen Siegerzelle des Wettbewerbsprozesses nicht allmählich(!) ansteigen, und es kann nicht ein Höchstwert dadurch erreicht und die kompetitive Phase durch dieses Indiz abgeschlossen werden. Die Feuerfrequenz kann allerdings – durch die Wirksamkeit hemmender Verbindungen – herabgesetzt werden. Wenn wir an der Auffassung festhalten, dass die Ausgabeschicht aus erregenden Zellen besteht, setzt die gegenseitige Hemmung dieser Zellen biologisch die Einführung eigener hemmender Zellen voraus, was kein besonderes Problem darstellt.

Wenn man diese besonderen Erfordernisse beachtet, kann tatsächlich eine Reduktion der Zahl der erregten Zellen nach Art des Wettbewerbsprozesses in Winner-take-all-Systemen erzielt werden.

In der Simulation mit der einfachen Anordnung der Abbildung 2.3.2–2 reduziert sich die Zahl der feuernenden Zellen von vier auf zwei. Die oberste Zellschicht enthält die zusätzlich erforderlichen hemmenden Neuronen. Die Verbindungsgewichte zwischen Eingabe- und Ausgabeschicht sind nicht zufällig bestimmt, sondern einfach so gestaffelt, dass die Zellen von *i1* bis *i5* aufwärts stufenweise jeweils etwas stärkere Verbindungen bilden. Die Ausgabezelle *o5* ist also begünstigt. Die Hemmung ist so eingestellt, dass nicht

schon ein einzelner Impuls auf die hemmende Zelle diese zum Feuern bringt, sodass eine häufiger erregte Ausgabezelle die Konkurrenten stärker hemmt.



749

Abbildung 2.3.2-2: Simulation eines Winner-take-all-Prozesses in einer Architektur mit biologischen Neuronen. Dargestellt ist ein Zustand nach dem Erreichen des gewünschten Effekts. Die Zellen $i1$ und $i2$ sind in der Refraktärphase (durch Schraffur angedeutet) und haben die Zellen $o4$ und $o5$ zum Feuern gebracht. Die oberste Zellreihe besteht aus hemmenden Zellen, kenntlich an den mit n bezeichneten Verbindungen. Die Verbindungen zwischen der untersten und der mittleren Zellschicht sind abgestuft, die Symbolisierung in den die Verbindungsgewichte anzeigenden Kreisen ist nur teilweise noch erkennbar.

Simulation:

Winner-take-all-Prozess, Architekturdarstellung

Ergänzender Hinweis: Steuerung über die Leertaste ergibt den besten Eindruck vom Simulationsverlauf.

Die Abbildung 2.3.2-3 zeigt den Verlauf des Winner-take-all-Prozesses und das Einmünden in einen stabilen Zustand. Dass am Ende nicht eine einzige feuernde Zelle übrig bleibt, ist nicht kritisch.

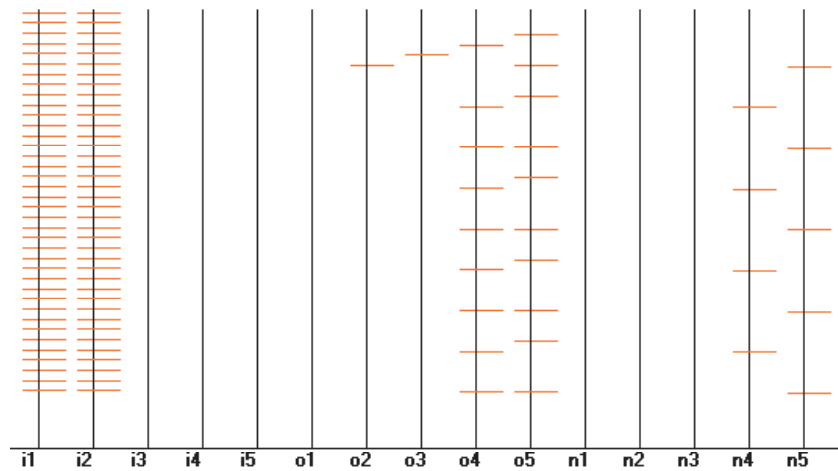


Abbildung 2.3.2-3: Verlauf eines Winner-take-all-Prozesses in der Architektur von Abbildung 2.3.2-2.

Simulation:
 Winner-take-all-Prozess, Verlauf
 Ergänzender Hinweis: Mit „Start-Simulation bis Stop“ entsteht sofort das komplette Verlaufsbild.

Wenn wir einmal zugestehen, dass ein Winner-take-all-Prozess prinzipiell möglich ist, bleibt doch das Problem des Zeitpunkts, an dem der Erfolg festgestellt werden soll. Man könnte sich ein übergeordnetes Netz vorstellen, das die Zahl der noch aktivierten Zellen der Ausgabeschicht *innerhalb einer bestimmten Zeitspanne* überwacht. Eine solche Konstruktion hätte aber angesichts der biologischen Realität wenig Wahrscheinlichkeit für sich.

Noch problematischer ist der Vorgang der Anpassung der Synapseneffektivität, der ja das eigentliche Ziel darstellt. Diese Anpassung sollte in Abhängigkeit von der Intensität der Inputaktivierung und der Zahl der aktivierten Zellen erfolgen, und zwar erst nachdem der Wettbewerbsprozess abgeschlossen ist! Es ist völlig offen, wie das mit biologischen Neuronen bewerkstelligt werden soll.

Das Vorbild des „competitive learning“ hat also Eigenschaften, die in einem biologischen System nicht möglich sind. Konsequenz ist, dass man offenbar schon in den Grundideen wenigstens teilweise anders verfahren muss.

Die hier vorzuschlagende Neukonstruktion hat folgende Eigenschaften:

Man kann vorläufig(!) die Vorstellung beibehalten, dass die beteiligten Zellen in eine Eingabeschicht und eine Ausgabeschicht aufgeteilt werden können. Auf die hemmenden Zellen wird verzichtet, da die Zahl der hemmenden Verbindungen bei komplexeren Modellen ohnehin zu stark anwächst (das hat bei standard-konnektionistischen Modellen zur Verwendung nicht-neuronaler Abkürzungsstrategien geführt). Die Anfangsgewichte der Verbindungen, die dem Lernprozess unterliegen, werden auf gleichmäßig kleine Werte gesetzt, die nach wie vor erforderliche Streuung des Lernerfolgs wird dadurch erzielt, dass die Lernbereitschaft der Ausgabezellen als Ganzes (hier „Plastizität“ genannt) gestreut ist. (Eine Streuung der Anfangsgewichte wäre auch möglich und würde zu entsprechenden brauchbaren Ergebnissen führen.) Für die Veränderung der Synapsengewichte im Lernprozess gilt, dass benutzte Verbindungen sofort verstärkt werden. Die wichtigste, die Funktion sichernde Annahme ist, dass eine Zelle, wenn sie feuert, ihre Lernbereitschaft (Plastizität) drastisch reduziert und auf diesem reduzierten Pegel eine mehr oder weniger lange Zeitspanne verbleibt. Die (ziemlich banale) Architektur ist in Abbildung 2.3.2–4 wiedergegeben. Man sieht, dass gegenüber Abbildung 2.3.2–2 einfach die hemmenden Zellen weggelassen sind.

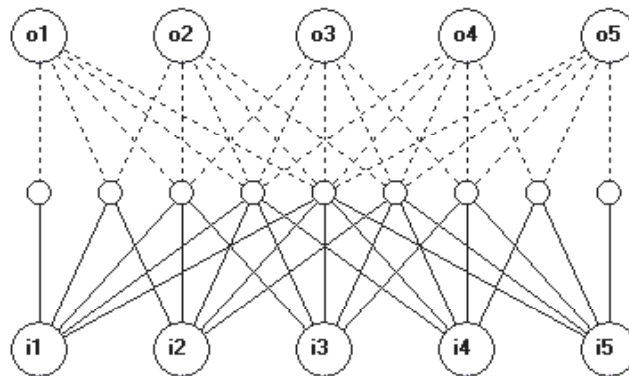


Abbildung 2.3.2–4: Architektur zur Simulation eines Lernprozesses mit biologischen Neuronen.

Der Input besteht in der Eingabe verschiedener Erregungsmuster: zunächst die Kombinationen $i1+i2$, $i2+i3+i5$ und $i2+i4$ mit einem Impulsburst, anschließend dieselben Kombinationen noch einmal mit einzelnen Impulsen. Die Abbildung 2.3.2–5 zeigt den Simulationsverlauf.

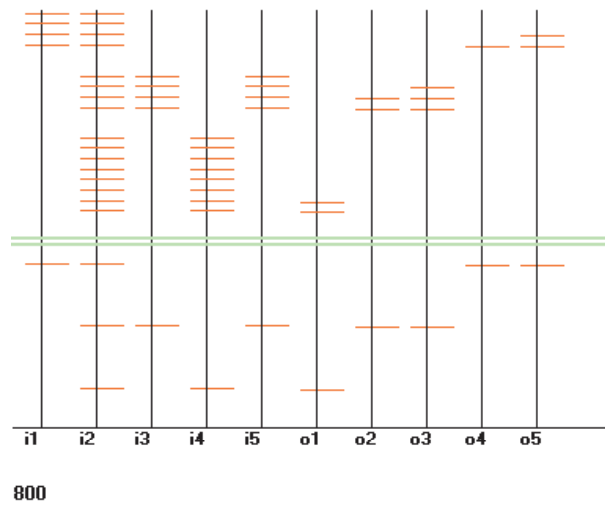


Abbildung 2.3.2–5: Verlauf des Lernprozesses in der Architektur von Abbildung 2.3.2–4. Unterhalb der Doppellinie die „Kannphase“.

Simulationen:
 Lernprozess, Architekturdarstellung
 Lernprozess, Verlauf; „Start–Simulation bis Stop“ führt zur „Kannphase“

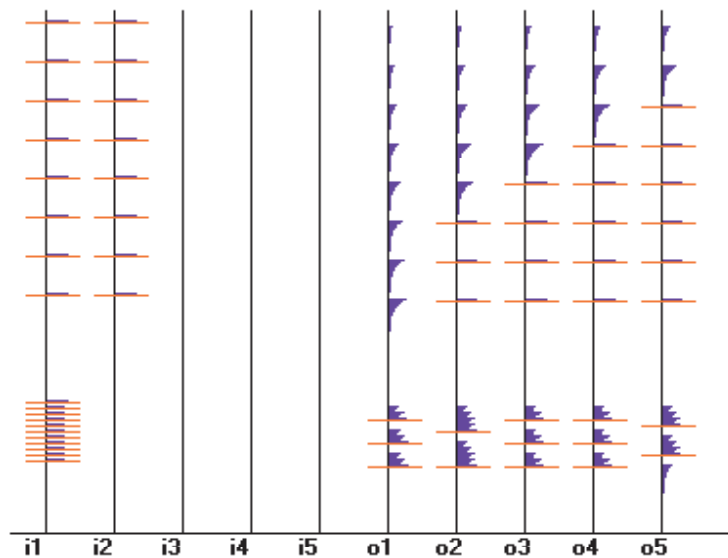
Das Ergebnis der Simulation zeigt, dass die Inputmuster je verschiedene Repräsentationen gebildet haben, und zwar nicht so, dass genau *eine* Zelle jeweils als Träger der Repräsentation dient, sondern mit einer gewissen Redundanz: Die zuerst eingegebenen Inputmuster werden durch jeweils zwei Zellen repräsentiert. Man beachte, dass es nicht die Kombination von zwei Zellen ist, sondern dass jede einzelne dieser Zellen für sich genommen vollständig gleichberechtigt die Repräsentationsaufgabe wahrnimmt.

Die Grundgedanken, die hinter dem beschriebenen Lernverfahren stehen, können kurz wie folgt zusammengefasst werden:

- Konzepte (bzw. deren Repräsentation) können als Koinzidenzdetektoren verstanden werden.
- Sie entstehen durch Verstärken der entsprechenden (aktivierten) Verbindungen auf einzelnen Neuronen, ...
- ... solange ein betroffenes Neuron noch nicht als Koinzidenzdetektor funktionsfähig ist.

- Die Funktionsfähigkeit wird mit dem Feuern des betroffenen Neurons erreicht.

Es ist nun von großer Bedeutung, dass das beschriebene Verfahren nur unter einer weiteren Bedingung funktioniert, die sich aus dem folgenden Simulationsexperiment ergibt. Die verwendete Architektur ist identisch mit der von Abbildung 2.3.2–4. Der Input ist dahingehend verändert, dass anschließend an eine Kombination von zwei Zellen ein „Muster“ bestehend aus der Aktivierung einer einzelnen Zelle mit erhöhter Frequenz eingegeben wird. Es sollte nun eigentlich das Ergebnis entstehen, dass die Aktivierung der einzelnen Eingabezelle auch noch zu einer brauchbaren Repräsentation auf der Ausgabeschicht führt, was aber nicht der Fall ist. Den Simulationsverlauf zeigt die Abbildung 2.3.2–6.



267

Abbildung 2.3.2–6: Verlauf des Lernprozesses in der Architektur von Abbildung 2.3.2–4 mit problematischer Frequenzerhöhung. Im Unterschied zu Abbildung 2.3.2–5 sind hier nicht nur die Aktionspotenziale, sondern auch die unterschwelligen Membranpotenziale wiedergegeben. Ein (blauer) Strich nach rechts von der senkrechten Mittellinie aus symbolisiert ein exzitatorisches Potenzial. Man sieht die exponentielle Abnahme der Potenziale und die Summation bei entsprechend rascher Aufeinanderfolge der Inputimpulse.

Simulationen:
 Frequenzproblematik, Architekturdarstellung
 Frequenzproblematik, Verlauf

Das Simulationsergebnis ist insofern enttäuschend, als auf den Input hin *alle* Zellen der Ausgabeschicht feuern, auch die dem konkurrierenden Muster schon zugeordneten. Es ist also nicht nur der Lernprozess selbst gestört, sondern auch die Aktivierung der gelernten Repräsentationen funktioniert nicht. Der Grund liegt darin, dass bei ausreichend rascher Aufeinanderfolge der Inputimpulse die ausgelösten EPSPs sich summieren und dadurch ggf. immer alle von einer Eingabezelle aus erreichbaren Ausgabezellen überschwellig erregt werden.

Damit ist auch die Lösung des Problems vorgegeben: Die Bildung lokalistischer Repräsentationen in biologischen Architekturen setzt einen ausreichenden Abstand der Aktionspotenziale voraus. Ausreichend heißt: so, dass eine unerwünschte Summation der EPSPs ausgeschlossen bleibt. Damit muss der Abstand größer sein, als das Zeitfenster, in dem synaptische Potenziale bei den betroffenen Zellen summiert werden. Das wiederum heißt, dass eine aktivierte Eingabezelle nur einen einzigen Impuls zur Bildung einer komplexeren Repräsentation beisteuern kann. Diese Voraussetzung bestimmt die Formierung von Repräsentationen grundsätzlich, und wir bezeichnen sie im Folgenden als „Einzelimpulskodierung“. Es ist in früheren Veröffentlichungen (Kochendörfer, 2000 und 2002) darauf hingewiesen worden, dass die im Kortex beobachtbaren Frequenzen an einzelnen Zellen, verglichen mit dem raschen Ablauf der sprachverarbeitenden Prozesse, schon für sich genommen die Annahme der Einzelimpulskodierung nahe legen. Ähnliche Überlegungen finden sich explizit auch bei Rieke et al. (1997: 55).

Das sich insgesamt ergebende Lernverfahren weicht in interessanter Weise von Hebb'schem Lernen in der für „spikende“ Neuronen und in der biologischen Diskussion üblichen Form ab.

Die immer wieder zitierte Fassung von Hebb lautet:

„When an axon of cell *A* is near enough to excite cell *B* or repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that *A*'s efficiency, as one of the cells firing *B*, is increased.“ (Hebb, 1949: 62)

In der Interpretation für eine Repräsentation mit „spikenden“ Neuronen wird daraus z. B.:

„A synaptic change occurs if presynaptic spike arrival and postsynaptic firing coincide within some time window;[...]“ (Gerstner et al., 1999:356)

Das bedeutet, dass die am Feuern beteiligten Synapsen verstärkt werden und Synapsen, die nicht beteiligt sind, unverändert bleiben oder je nach den verwendeten Strategiedetails auch abgeschwächt werden. Um ein andauerndes Anwachsen der Synapseneffektivität zu verhindern, wird in technischen Systemen meist eine Obergrenze für die Verstärkung festgelegt, in biologischen Neuronen versteht sie sich von selbst.

Die wichtigsten Unterschiede des oben vorgeschlagenen Verfahrens gegenüber dem Hebb'schen Lernen bestehen in den folgenden Punkten:

- Die Verstärkung von Synapsen ist unabhängig vom Feuern des postsynaptischen Neurons.
- Das Feuern des postsynaptischen Neurons bewirkt einen lokal bestimmten Abschluss des Lernprozesses.
- Der Lernprozess kommt ohne Spontanaktivität zustande.
- Es gibt keine besondere Funktion einer Obergrenze für die Leistungsfähigkeit von Synapsen.

Da das Hebb'sche Lernen die Vorstellungen von Lernprozessen in biologischer Literatur stark beeinflusst, lohnt sich die Überlegung, ob die Differenzen wirklich unvermeidbar sind bzw. ob nicht doch die für spikende Neuronen abgeleitete Version einfach übernommen werden kann.

Eine für das Hebb'sche Lernen kritische Situation entsteht dann, wenn zwei Eingabemuster Komponenten gemeinsam haben, wie in Abbildung 2.3.2–7 schemantisch dargestellt.

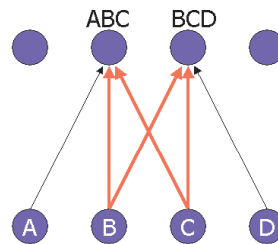


Abbildung 2.3.2–7: Schema zum Hebb'schen Lernen von Repräsentationen, die gemeinsame Komponenten haben.

Wenn in diesem Beispiel versucht wird, zwei Eingabemuster $A+B+C$ und $B+C+D$ zu repräsentieren, dann ist das nur unter unbequemen Zusatzbedingungen möglich. Zunächst für eine Lernregel ohne Abschwächung nicht beteiligter Synapsen: Da für die Synapsengewichte eine Obergrenze gilt und diese Obergrenze bei genügend vielen Lernschritten immer erreicht wird, ist es von dieser Obergrenze abhängig, ob $A+B+C$ und $B+C+D$ noch unterschieden werden können. Wenn die Obergrenze so ist, dass zwei Eingänge ausreichen, um ein Neuron zum Feuern zu bringen, sind $A+B+C$ und $B+C+D$ identisch, das heißt, beide Repräsentationen sind immer gleichzeitig aktiviert. Dasselbe gilt für alle weiteren Repräsentationen, die $B+C$ enthalten. Wenn man niedrigere Obergrenzen annimmt, entsteht das Problem analog mit mehr als zwei Komponenten. Wenn man nicht beteiligte Synapsen abschwächt, entstehen zusätzlich Interferenzen: Das Lernen von $B+C+D$ stört die Repräsentation von $A+B+C$, ohne dass das Differenzierungsproblem gelöst würde.

Dieses Differenzierungsproblem kann für das Verständnis linguistischer Strukturen und Prozesse nicht akzeptiert werden. Mit der Überlappung von Konzeptkomponenten ist in großem Umfang zu rechnen. Eine Lösung des Problems kann nur darin bestehen, dass der Prozess der Verstärkung von Synapsen zu einem geeigneten Zeitpunkt abgebrochen wird. Dieser Zeitpunkt kann nicht für alle Fälle fest vorbestimmt werden, und es gibt wohl kaum eine andere Möglichkeit als die, das Feuern der repräsentierenden Zelle dafür heranzuziehen. Damit kann das Feuern nicht mehr als Bedingung für die Verstärkung von Synapsen verstanden werden.

Es wird nun allerdings in experimentellen Studien, hauptsächlich mit Bezug auf den Hippokampus, festgestellt, dass eine Langzeitpotenzierung (LTP) nur dann zu beobachten ist, wenn ein synaptischer Input in eine Zelle mit dem Feuern dieser Zelle koordiniert ist (im Sinne von Hebb; vgl. Bi & Poo, 2001). Nun kann zwar der Hippokampus wegen seiner besonderen neuronalen Struktur nicht ohne weiteres stellvertretend für den Kortex als Ganzes stehen, und es ist auch von vornherein damit zu rechnen, dass mehrere verschiedene Lernprozeduren für die Funktion des Kortex erforderlich sind. Auf der anderen Seite führt die Idee der synaptischen Verstärkung ohne Voraussetzung des Feuerns doch auch noch zu einer störenden Nebenwirkung, die darin besteht, dass vielfach Synapsen verstärkt werden, ohne dass sich daraus ein sinnvoller Lerneffekt ergibt (also ohne dass die betroffenen Zellen zum Feuern kommen). Diese Synapsenverstärkung muss sich störend bei nachfolgenden Lernprozessen bemerkbar machen.

Prinzipiell ist dieses Problem dadurch zu lösen, dass Synapsen durch einen Vergessensprozess wieder geschwächt werden. Wünschenswert ist hier aber

eine Asymmetrie, die einen relativ raschen Vergessensprozess bei „ungültig“ verstärkten Synapsen und einen viel langsameren Vergessensprozess bei „gültiger“, zur Bildung einer Repräsentation führender Verstärkung vorsieht. Es ist – jedenfalls im Modell – relativ einfach, diese Asymmetrie zu erzielen, indem man nämlich von einem zunächst raschen Vergessensprozess ausgeht und beim Feuern der Zellen, also beim Abschluss des Lernprozesses, nicht nur die weitere Verstärkung, sondern auch den Abbau der Synapseneffektivität bremst.

Ein Nebeneffekt dieser Maßnahme ist dann, dass, abhängig von den konkret gewählten Parametern, tatsächlich eine Langzeitpotenzierung, das heißt jetzt: ein längeres Andauern der Synapsenverstärkung, nur erreicht wird, wenn die Zelle feuert. Das ist nicht mehr die Idee Hebbs, kann aber die vorliegenden Beobachtungen erklären.

Wir können abschließend festhalten, dass eine unter dem Aspekt möglicher Lernprozesse den Ansprüchen genügende Repräsentationsform lokalistisch sein muss, mit zwei ergänzenden Zusatzbedingungen: Es muss Einzelimpulskodierung angenommen werden, und die Lernprozesse selbst müssen abweichend von der Hebb'schen Idee strukturiert werden, ohne dass dabei notwendig die der Hebb'schen Idee entsprechenden Beobachtungen angefochten werden müssen.

2.3.3 Konzeptbildung

Im vorangegangenen Abschnitt sind elementare Repräsentationsvorgänge behandelt worden, mit zunehmender Beschränkung auf biologische Systeme. Wenn im Folgenden von Konzepten die Rede ist, werden damit ausdrücklich *mentale* Repräsentationen angesprochen und der Schwerpunkt liegt nicht so sehr bei elementaren Lernprozessen, sondern beim Zustandekommen komplexerer Gebilde. Konzepte müssen in einem lokalistischen System apparative Realitäten (potenzielle Repräsentationen) sein.

In der strukturalistischen Tradition der Linguistik wird „Konzept“ für die Inhaltsseite eines Zeichens verwendet. Demgegenüber ist hier der Begriff wesentlich erweitert. Man kann seinen Umfang dadurch bestimmen, dass man den Zusammenhang mit der Kategorisierung schlechthin beachtet. Konzepte leisten Kategorisierung, was allerdings nicht ihre einzige Funktion ist. Damit sind, um bei linguistischen Beispielen zu bleiben, auch Phoneme und syntaktische Kategorien Konzepte.

Konzepte sind Einheiten. Sie werden in lokalistischen Architekturen durch einen einzelnen Knoten repräsentiert. Ein solcher Knoten ist aus verarbei-

tungstechnischen Gründen, wie später zu zeigen sein wird, nicht identisch mit einer einzelnen Nervenzelle. Man kann aber vorläufig diese Vorstellung beibehalten und die Bezeichnung „Großmutterzelle“ – mit einer entsprechenden Warntafel versehen – für eine Konzeptrepräsentation verwenden.

In den zweischichtigen Modellen des vorigen Abschnitts waren die Zellen der Outputschicht zunächst nicht einem Inputmuster zugeordnet und sie waren in diesem Sinne bedeutungslos. Mit dem Entstehen der Zuordnung im Lernprozess haben sie eine spezielle Funktion, also Bedeutung erhalten. Die Zellen der Inputschicht, die ja auch Großmutterzellen sind, waren von vornherein mit einer Bedeutung versehen zu denken, und es ergibt sich die Frage, wie diese Zellen zu ihrer Bedeutung kommen. Diese Frage stellt sich in verschärfter Form, wenn man nicht mit einer zweischichtigen Architektur rechnet, sondern mit einer freieren Anordnung. Es sind vielfach „Modelle“ konstruiert worden, in denen einfach Kreise (für Knoten bzw. Zellen) mit Etiketten versehen worden sind. Ein Beispiel gibt Abbildung 2.3.3–1. Damit bleibt natürlich die Frage, woher die etikettierten Zellen wissen, welches Etikett sie tragen (um im Bild zu bleiben), ungelöst.

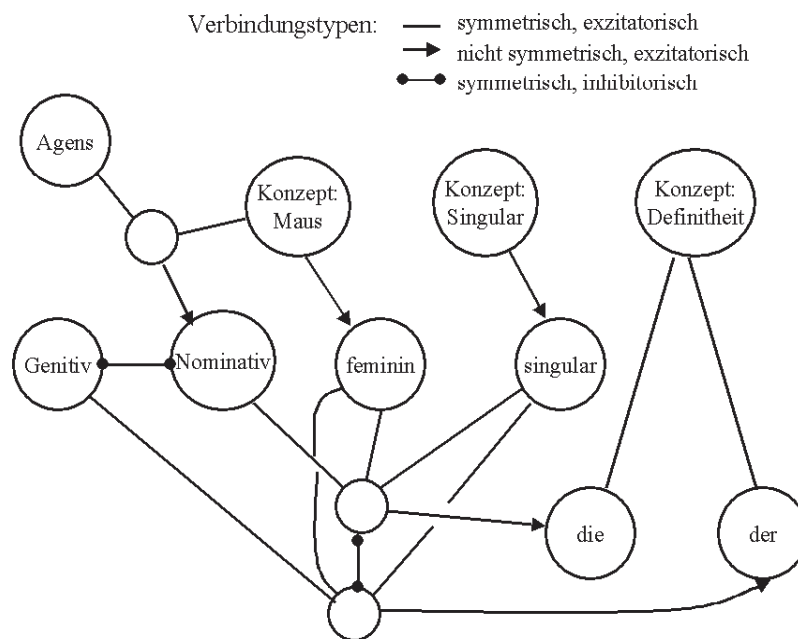


Abbildung 2.3.3–1: Lokalistische Einheiten in einem einfachen Produktionsmodell für Wörter. (Verändert nach Schade, 1992: 90.)

Wenn Zellen über die Aktivierung anderer Zellen durch Lernprozesse eine Bedeutung bekommen können, kann man natürlich diese nun bedeutungshaltigen Zellen ihrerseits in Lernprozesse einbeziehen und weitere Zellen mit Bedeutungen ausstatten. Aber auf irgendeiner Ebene muss dieser fortgesetzte Lernprozess seinen Ausgangspunkt haben. Es muss als Basis aller Konzeptbildungsprozesse eine Menge von Primitiven geben, das heißt eine Menge von Großmutterzellen, die immer schon eine Bedeutung tragen. Es ist, schon allein um eine Anbindung der gelernten bzw. zu lernenden Konzepte an die Realität der Umwelt zu haben, und solange wir an Perzeptionsvorgänge denken, völlig klar, dass diese Funktion von den Sinneszellen und den damit unmittelbar und ohne Lernvorgänge verbundenen neuronalen Strukturen wahrgenommen wird. Beispiele für Großmuttereinheiten an der Sinnesperipherie sind oben in Abschnitt 2.3.1 aufgeführt.

An dieser Stelle sei ergänzend darauf hingewiesen, dass – unter der Voraussetzung, dass Lernprozesse an Einzelimpulskodierung gebunden sind und die Sinnesperipherie Erregungsintensitäten in Impulsfrequenzen kodiert – ein Übersetzungsprozess in eine entsprechende Einzelimpulskodierung vorausgesetzt werden muss. Details werden in Teil 3 am Beispiel der Phonetik bzw. Phonologie besprochen.

Es ist nun zu überlegen, wie Konzeptbildungsprozesse auf der Basis von Sinnesprimitiven im Detail aussehen. Die einfachste Annahme wäre vielleicht die, dass es sich tatsächlich um einen zweistufigen Prozess handelt, bei dem Mustern von Sinnesprimitiven in *einem* Schritt Konzepte zugeordnet werden. Dies würde einer ganzheitlichen (holistischen) Kategorisierung von Erscheinungen der Wirklichkeit entsprechen.

In Kochendörfer (2002: Kap. 2.4) ist darauf hingewiesen worden, dass es Ressourcenprobleme geben kann, wenn Konzepte aufgrund einer größeren Anzahl von Komponenten gebildet werden sollen, denn es muss eine entsprechend große Anzahl von passenden verstärkbaren Verbindungen auf den einzelnen Zellen vorausgesetzt werden, die dann als Großmutterzellen dienen sollen.

Ein weiteres Argument ergibt sich, wenn man einfache Beobachtungen heranzieht, die etwas mit Konzeptbildungsprozessen zu tun haben. Wenn man ein unbekanntes Objekt sieht, das man auch nicht benennen kann, ist man doch in der Lage, Eigenschaften anzugeben: „Ist rot, so lang wie ein Unterarm, insgesamt eher flach, hat längliche Vertiefungen...“. Die angegebenen Eigenschaften (die durch die gebrauchten Wörter zitiert werden) sind selbst Konzepte, die nicht als elementar angesehen werden können, sondern vorangegangene Lernprozesse voraussetzen. Welche Rolle spielen solche Konzepte bei der Konstitution des Konzepts, das dem bis dahin unbekanntem Objekt

entspricht? Bei der Annahme einer einschrittigen Konzeptbildung lautet die Antwort auf diese Frage: keine. Alle für die Konzeptbildung herangezogenen Komponenten sind als elementar zu betrachten. Das bedeutet aber, dass jedes zu etablierende Konzept vollkommen neu ist, und damit auch, dass die eventuelle Vertrautheit mit Komponenten keinen Einfluss auf den Grad der Lernbarkeit hat. Das widerspricht entschieden der alltäglichen Erfahrung.

Das Fazit aus diesen Überlegungen ist, dass die Annahme lokalistisch repräsentierter Konzepte im Kortex praktisch immer auch bedeutet, dass die Konzeptbildung mehrschrittig erfolgen muss. Damit kann die Verbindung mit der Sinnesperipherie mehr oder weniger indirekt werden, das heißt: vermittelt durch beliebige Gedächtnisinhalte, die z. B. auch prozeduralen Charakter haben können.

Auf einige weitere Details zur Konzeptbildung, die sich bei genauerer Analyse der Simulationsexperimente des vorigen Abschnitts ergeben, soll hier nur stichwortartig hingewiesen werden:

- Wenn Inputmuster häufiger auftreten, werden mehr gleichartige Konzeptknoten rekrutiert als bei selteneren Inputmustern. Diese Form von Redundanz ist nicht durch irgendeine prinzipielle Maßnahme vermeidbar.
- Wenn Inputmuster auf ein „leeres“ Gehirn treffen, werden mehr Konzeptknoten rekrutiert, als bei einer bereits stark „besetzten“ Ausgangsstruktur (Begrenzung der Redundanz).
- Es kann nicht verhindert werden, dass *Teile* eines Inputmusters ihrerseits Konzepte bilden. Das kann die Grundlage für ein Lernen aus Beispielen sein.

Eine wichtige Folge der mehrschrittigen Konzeptbildung ist die Filterfunktion der Komponenten. Wahrgenommen wird nur, wofür wir Organe haben. Elementarere Komponenten, die für eine Konzeptbildung verfügbar sind (und nicht neu aufgebaut werden müssen) bestimmen, was beachtet wird. Wegen der Frequenzabhängigkeit der Lernprozesse beim Aufbau dieser elementareren Komponenten ist dies gleichbedeutend mit einer Abstraktion, die das „Wesentliche“ heraushebt. Man kann sich ausmalen, dass diese Abstraktion in der Evolution durch ihre überlebenssichernde Funktion begünstigt worden ist.

2.3.4 Konsequenzen für das Verständnis kortikaler Strukturen

In diesem Abschnitt sollen einige Konsequenzen behandelt werden, die sich aus der Argumentation in den vorangegangenen beiden Abschnitten sofort ergeben, vor allem solche, die gegenüber weithin akzeptierten Vorstellungen in der Biologie abweichen und damit zur Kritik Anlass geben können. Ein vollständigeres Bild (z. B. die Differenzierung in verschiedene Zelltypen einschließend) kann erst später entworfen werden.

Wenn der Kortex senkrecht zur Oberfläche geschnitten wird, zeigt sich eine deutliche Schichtung. Diese Schichtung entsteht dadurch, dass Zelltypen und Dichte der Zellkörper schichtartig verteilt sind (vgl. oben Abschnitt 2.1.2). Es ist aber nicht so, dass innerhalb einer Schicht sich die gleichen Bestandteile, wieder in Schichten organisiert, wiederholen. Da die Dendritenbäume der Pyramidenzellen eher näher der Kortexoberfläche liegen und die Axone hauptsächlich von dort aus gesehen nach unten gerichtet sind, kann man vereinfacht sagen, dass die Haupt-Verarbeitungsrichtung von oben nach unten verläuft. Bestandteile einer Pyramidenzelle können auf alle Schichten verteilt sein. Diese Beobachtungen haben, zusammen mit dem Erfordernis einer mehrschrittigen Konzeptbildung, die Folge, dass man annehmen muss, dass Verarbeitungsprozesse den Kortex mehrfach in der Senkrechten durchlaufen können. Die dafür erforderlichen von unten nach oben orientierten Verbindungen bleiben teilweise innerhalb des Kortex oder bestehen in myelinisierten Fasern, die den Kortex an einer Stelle verlassen und an einer anderen Stelle wieder in den Kortex hineinführen. Die diesbezüglichen Möglichkeiten sind in Abbildung 2.3.4–1 schematisch zusammengestellt.

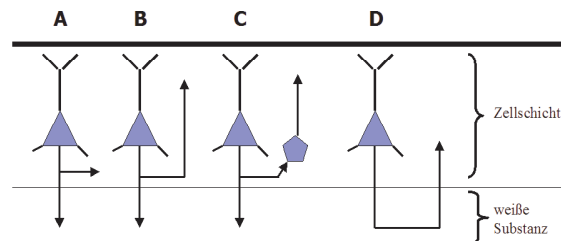


Abbildung 2.3.4–1: Beispiele für Verbindungstypen im Kortex, die eine Weiterverarbeitung von Verarbeitungsergebnissen ermöglichen. **A.** Verbindung mit basalen Dendriten anderer Pyramidenzellen. **B.** Verbindung mit apikalen Dendriten anderer Pyramidenzellen. **C.** Verbindung vermittelt über eine Sternzelle (Martinotti-Zelle). **D.** myelinisierte Verbindung, die über die weiße Substanz geht.

Der Signalfluss kann dann als mäandrierend gedacht werden, wie in Abbildung 2.3.4-2 veranschaulicht.

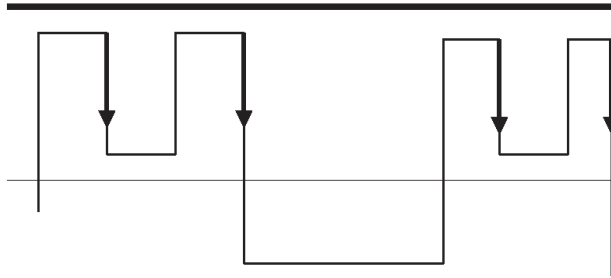


Abbildung 2.3.4-2: Mäandrierende Verarbeitung z. B. bei der Kategorisierung von Input. Aktivierte Bahnen, die später nicht weiterführen, müssen an jeder Pfeilspitze hinzugedacht werden.

Die Konsequenz ist schließlich, dass auch ein einfacher Verarbeitungsprozess nicht nur eine punktförmige Stelle des Kortex betrifft, sondern dass immer eine flächenhafte Anordnung von Zellen eine sich rasch ausbreitende Aktivierung zeigen wird, wie in 2.3.4-3 dargestellt. Das ist wichtig für die Interpretation von Ergebnissen „bildgebender Verfahren“.

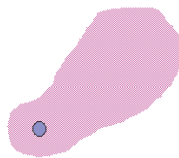


Abbildung 2.3.4-3: Bei Draufsicht auf die Kortexoberfläche: Fläche aktivierter Neuronen, schematisch, nach Input an der Stelle des Kreises.

Flächenhafte Aktivierungen entstehen aber auch schon dadurch, dass der Lernprozess in einer lokalistischen Architektur immer ein Überangebot an verstärkbaren Verbindungen voraussetzt, die alle versuchsweise aktiviert werden müssen, und von denen dann nur einige tatsächlich verstärkt werden. Diese Voraussetzungen sind in Abbildung 2.3.4-4 noch einmal dargestellt.

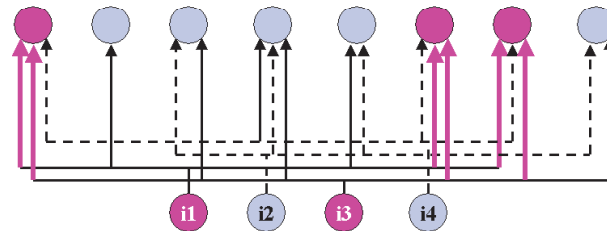


Abbildung 2.3.4–4: Überangebot verstärkbarer Verbindungen. Nicht benutzte Verbindungen sind gestrichelt, die durch das Eingabemuster $i1+i3$ erregten Verbindungen sind durchgezogen, nur die fett (bzw. rot) ausgezeichneten Verbindungen werden tatsächlich dauerhaft verstärkt.

Es ist auch sofort klar, dass nicht nur ein Überangebot von Verbindungen sondern auch ein Überangebot von Zellen vorhanden sein muss. Wenn es möglich ist, dass vorhandene Verbindungen nicht benutzt werden (nicht verstärkt werden), und damit bestimmte Zellen nicht überschwellig erregt werden können, gilt banalerweise, dass diese Zellen zu einem Vorrat gehören, der ggf. zu einem späteren Zeitpunkt benutzt wird. Wenn von einem Überangebot von Verbindungen die Rede ist, sollte man beachten, dass nicht an eine vollständige Verbindung z. B. einer Zellschicht mit einer zweiten zu denken ist, geschweige denn an eine Verbindung aller Zellen des Kortex miteinander. Daraus folgt, dass man damit rechnen muss, dass es Zellen gibt, die über einen großen Zeitraum oder lebenslang gar nicht zur Konzeptrepräsentation verwendet werden, weil die für sie möglichen Verbindungskombinationen im Input nicht vorkommen. Es kann auch eine Konsequenz von Lernprozessen sein, dass bestimmte Zellen dauerhaft nicht mehr aktiviert werden können. Lernprozesse können konkurrierende Lernvorgänge und damit die entsprechenden Zellpopulationen ausschließen.

Wenn man unter diesen Voraussetzungen die tatsächlich beobachtbaren neuronalen Strukturen betrachtet, wird deutlich, dass die hohe Konnektivität im Kortex (größenordnungsmäßig 10 000 Verbindungen pro Zelle) eine Voraussetzung für Lernprozesse ist. Es ist auf diese Weise auch erklärbar, dass die überwiegende Anzahl der Verbindungen sehr schwach ist. Allerdings wird gefordert, dass es herausgehobene, verstärkte Verbindungen gibt.

Man beachte in diesem Zusammenhang die Überlegungen bei Braitenberg & Schüz (1998: 63):

„Eight thousand synapses on each neuron are too many if we are to see neurons as „threshold devices“, as units which respond „all or none“ to specific constellations of activity in their input lines. If the threshold is to be set exactly at some number between 0 and 8000 in order to define the set (or sets) of active input lines to which the neuron is supposed to respond, we have to assume an immaculate precision of the electrochemical phenomena in the neuron membrane which no physiologist would deem possible.“

Das heißt, wenn die Aktivierung oder Nicht-Aktivierung einzelner Verbindungen wirksam sein soll, muss eine unrealistisch genaue Einstellung des Schwellenwerts angenommen werden – unter der Voraussetzung, dass die betreffenden aktivierten Verbindungen nicht von einer entsprechenden Stärke sind.

Dieses Argument wird verschärft, wenn man beachtet, dass die Aktivität der Inputverbindungen zeitlich gestreut ist. Man muss damit rechnen, dass zwischen zwei Inputimpulsen Zeitspannen im Millisekundenbereich liegen. In dieser Zeit klingt das EPSP der Zelle, die diese Impulse erhält, entsprechend ab. Wenn diese Zeitspannen nicht exakt vorhersagbar sind und wenn der Input sich aus sehr vielen schwachen Synapsen ergibt, ist das Verhalten auch dann praktisch unabhängig von der Aktivität eines einzelnen Eingangs, wenn die Schwelle (unrealistisch) exakt eingestellt werden kann. (Man vgl. das Experiment in Kochendörfer, 1997: 61 ff.)

Die Konsequenz ist: Verstärkte Verbindungen müssen *deutlich* verstärkt sein und sich damit deutlich von einem Hintergrund u. U. ebenfalls aktivierter schwacher Verbindungen abheben. Eine verstärkte Verbindung in diesem Sinne muss allerdings nicht durch eine einzelne Synapse gebildet werden, es ist auch möglich, dass ein Synapsenbündel, das von einem einzelnen Axon ausgeht, dies leistet.

Am Rande sei hier auf eine Querbeziehung zum Problem der mehrstufigen Konzeptbildung hingewiesen. Wenn es schwierig ist, sich eine einigermaßen zuverlässige Verrechnung vieler schwacher Eingänge in eine Zelle vorzustellen, muss auch aus diesem Grund mit einer Verteilung der Verknüpfungen, die zu einem komplexen Konzept gehören, auf eine Hierarchie aus mehreren Zellen gerechnet werden (Abbildung 2.3.4-5).

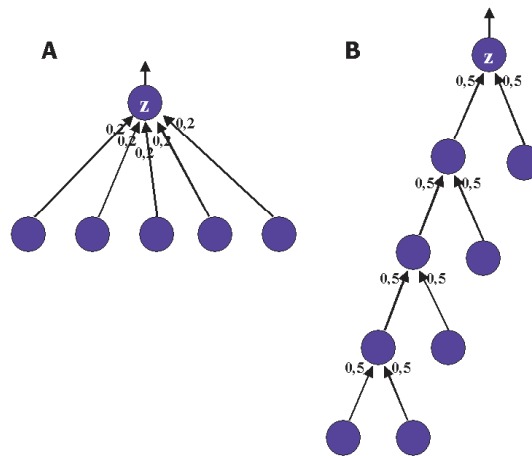


Abbildung 2.3.4–5: Synapsengewichte bei einstufiger und mehrstufiger Konzeptbildung. Die Konstruktion in B kommt mit wesentlich größeren Einstellungen der Synapsengewichte aus, der Effekt in den mit z bezeichneten Zellen ist, abgesehen vom zeitlichen Verlauf, derselbe.

Die Tatsache, dass die Zahl der Synapsen auf einer Zelle so hoch ist, ist also in dem hier interessierenden Bereich nicht zu erklären durch die Schwäche der einzelnen Synapse, was zur Folge hätte, dass viele Synapsen gebraucht würden, um eine nachgeschaltete Zelle überschwellig zu erregen, sondern es handelt sich um ein Angebot zunächst schwacher, aber verstärkbarer Synapsen, das es ermöglicht, die für das Funktionieren der Datenverarbeitung im Gehirn erforderlichen spezifischen Verbindungen aufzubauen.

Das Gehirn besteht in einer Mischung von Ordnung (mit spezifischer Funktion) und Chaos. Für den Kortex speziell gilt, dass, wenn man das Verbindungsmuster betrachtet, das Chaos überwiegt. So etwas wie eine gezielte und angeborene „Verdrahtung“ gibt es vor allem im subkortikalen Bereich und teilweise noch dort, wo größere Kortex-Areale in einer bestimmten Weise miteinander verknüpft sind. Anders dort wo Lernen stattfindet: Hier werden offenbar zwar Grundstrukturen verwendet, die verschiedene Zelltypen einbeziehen, greifbar in der Schichtung des Kortex. Weitere Verbindungen sind aber chaotisch. Man überlege, welche Möglichkeiten es prinzipiell geben könnte, eine bessere, geordnete Struktur zu rechtfertigen. Ein chaotisches Verbindungsmuster ist jedenfalls ideal geeignet, um als Basis für lokalistische Lernprozesse zu dienen.

Eine wichtige Frage ist die nach der Rolle von Spontanaktivität im Kortex. Die klassischen Erklärungen von Langzeitpotenzierung und Langzeitdepres-

sion setzen die Annahme von Spontanaktivität voraus, das heißt ein (spontanes) Feuern der beteiligten nachgeschalteten Zellen, das nicht durch einen Input verursacht ist, also nicht etwa ein Feuern, dessen Vorgeschichte man aus irgendwelchen Gründen nicht kennt.

Es gibt ein vergleichsweise einfaches Argument, das zeigt, dass es eine solche Spontanaktivität in dem benötigten Ausmaß und dort, wo mit Lernprozessen zu rechnen ist, nicht geben kann. Vorausgesetzt ist, dass Konzepte mehrstufig gebildet werden und dass also ein Berechnungsergebnis des Kortex wieder in den Kortex zur Weiterverarbeitung eingegeben werden kann. Dann sind nämlich alle Spontanaktivitäten (oder mindestens die von Pyramidenzellen) von Berechnungsergebnissen nicht zu unterscheiden und es finden automatisch Lernprozesse auf der nächsten Berechnungsstufe statt. Diese Lernprozesse würden aber dazu führen, dass der Kortex mit „fiktiven“ Konzepten angefüllt würde und sinnvolle Lernprozesse dadurch blockiert wären.

Das Funktionieren von Lernprozessen ist nur denkbar *ohne* Spontanaktivität im Kortex. Das ist eine wichtige zusätzliche Stütze der oben in Abschnitt 2.3.2 vorgeschlagenen Lernprozedur, die ohne Spontanaktivität auskommt. Nur ergänzend sei noch darauf hingewiesen, dass Sprachproduktionsprozesse unter der Annahme von Einzelimpulskodierung nicht in der beobachtbaren großen Zuverlässigkeit möglich wären, wenn es nennenswerte Raten von Spontanaktivität gäbe. Für die Einzelimpulskodierung sprechen aber, wieder von den Beobachtungen her argumentiert, die niedrigen Impulsfrequenzen, die vom Kortex abgeleitet werden.

2.3.5 Diskussion

Dieser Abschnitt dient der ergänzenden Diskussion von Argumenten, die *gegen* die Großmutterzellentheorie, also gegen eine lokalistische Auffassung von der Repräsentation im Gehirn angeführt worden sind und die von vielen Autoren als entscheidend angesehen werden.

Das am weitesten verbreitete Argument wird von Lyre (2002: 151 f.) so formuliert:

„...(die Unplausibilität so genannter „Großmutter-Neuronen“ ist bereits durch die damit verbundene Kostenexplosion einsehbar: die Zahl der mir empirisch zugänglichen Objekte überschreitet sichtlich die Zahl der hierfür potentiell zur Verfügung stehenden Neuronen).“

Der kritische Punkt liegt im Stichwort „empirisch zugänglich“: Man beachte: Empirisch zugänglich ist nur das, was in das Nervensystem passt und nur so viel als das Nervensystem zulässt, vgl. oben Abschnitt 2.3.3. Das Argument ist nur gültig, wenn man annimmt, dass beliebig große Mengen von Inputmerkmalen durch eine Großmutterzelle verknüpft werden müssen. Das ist aber nicht der Fall. Information wird im Wahrnehmungsprozess mehrfach gefiltert. Die erste Filterstation bilden die Sinneszellen. Eine weitere Ausdünnung geschieht durch die vorhandenen, durch Lernvorgänge aufgebauten elementarerer Konzepte.

Eine „Kostenexplosion“ soll auch durch das sog. „Bindungsproblem“ verursacht werden, das schon einmal oben in Abschnitt 2.2.4 angesprochen worden ist. Nach einer Idee von F. Rosenblatt, wiedergegeben aufgrund der Darstellung in Malsburg (1999): Wenn eine visuelle Darstellung oben ein Dreieck und unten ein Rechteck enthält, genügt es für den Wahrnehmungsprozess nicht, die Merkmale UNTEN, OBEN, RECHTECK und DREIECK zu aktivieren, sondern UNTEN muss an RECHTECK und OBEN an DREIECK gebunden werden. Diese Bindung ist in natürlicher Weise möglich, wenn man lokalistische Repräsentationen annimmt, sofern entsprechende Lernprozesse mit ausreichender Schnelligkeit plausibel gemacht werden können. Die Folge ist dann, wegen der vielen möglichen Kombinationen, ein dramatisches Anwachsen der Menge der erforderlichen Großmutterzellen. In diesem Zusammenhang muss aber beachtet werden, dass im Gehirn ständig Vergessensprozesse stattfinden. Man denke etwa an Kurzzeitgedächtnisphänomene, die Sekunden bis Minuten andauern, und die wenigstens teilweise kurzfristig wirksame Lernprozesse und ein anschließendes Vergessen voraussetzen. Wenn eine visuell wahrgenommene Form kurzfristig wiedererkannt werden kann und also über eine Großmutterzelle repräsentiert sein muss, bedeutet das somit nicht die dauerhafte Belegung dieser Zelle.

Die oben in Abschnitt 2.3.2 entwickelte These, dass es keine biologisch plausiblen Lernverfahren für verteilte Architekturen geben kann, ist deshalb interessant, weil gelegentlich genau das Gegenteil behauptet worden ist: Lokalistische Architekturen sollen Schwierigkeiten mit dem Lernen haben, nicht aber verteilte (vgl. Ziemke & Cardoso de Oliveira, 1996: 17). Richtig ist in der Tat, dass in der Vergangenheit eine Reihe von lokalistischen Sprachverarbeitungsmodellen – allen voran das Lexikonmodell TRACE (McClelland & Rumelhart, 1986) – konstruiert worden sind, zu denen keine Lernprozesse angegeben werden können. Der Grund dafür liegt in unbiologischen Architekturdetails, nicht in dem insgesamt lokalistischen Repräsentationsprinzip.

Die Verletzlichkeit lokalistischer Architekturen nach dem Slogan „Großmutterzelle weg, Großmutter weg“ ist wegen der unvermeidlichen Redundanz,

die durch die Lernprozesse verursacht wird, relativ gering einzuschätzen und Läsionen, die ja z. B. im Fall von Aphasien nachweislich zu Verlusten von Konzepten führen, müssen eine gewisse Ausdehnung erreichen, um solche Wirkungen zu haben. Die Störbarkeit des Zugriffs auf lexikalische Einheiten bei Aphasie ist offenbar von der Frequenz abhängig, die diesen lexikalischen Einheiten bezogen auf den Patienten zukommt. Die Bildung lokalistischer Repräsentationen (das heißt die Zahl der Großmutterzellen für ein bestimmtes Konzept) ist, wie die Simulationen in Abschnitt 2.3.2 zeigen, tatsächlich davon abhängig, wie häufig ein Input angeboten wird. Der Effekt ist insgesamt deutlicher und stabiler als bei verteilten Repräsentationen.

Man sollte auch einmal darüber nachdenken, dass, wenn eine Zelle in einer verteilten Repräsentation (Ausgabeschicht!) nicht mehr funktionsfähig ist, sich in jedem Fall die Repräsentation von Konzepten prinzipiell verändert hat. In einer lokalistischen Architektur ist das nicht der Fall, der Defekt kann über die Redundanz der Repräsentation ausgeglichen werden. Insgesamt dürften, was die Verletzlichkeit angeht, redundante lokalistische Architekturen gegenüber verteilten im Vorteil sein.

2.4 Semantik

2.4.1 Fragestellung

Wörterbücher sind alltäglich begegnende Produkte, die sich mit Semantik beschäftigen. Was man üblicherweise kennt, sind semasiologische Wörterbücher, das heißt Wörterbücher, die zu Wortformen Bedeutungen angeben. Weniger geläufig sind onomasiologische Wörterbücher, die angeben, wie man bestimmte Dinge und Sachverhalte benennt, die also, verglichen mit den semasiologischen Wörterbüchern, den umgekehrten Weg gehen.

Bei beiden Wörterbuchtypen wird letztlich die Beziehung zwischen Ausdrucksseite (in Saussureschen Termini „signifiant“) und Inhaltsseite („signifié“) thematisiert. In semasiologischen Wörterbüchern ist die Ausdrucksseite Ordnungskriterium. Typische Fragestellungen betreffen die Mehrdeutigkeit (Polysemie, Homonymie), im weitesten Sinne verstanden, der Ausdrucksseiten. Wenn man an zweisprachige Wörterbücher denkt, ist die Bedeutungsangabe durch ein bzw. mehrere Wörter der anderen Sprache ersetzt. Bei den onomasiologischen Wörterbüchern sind es die als gegeben geltenden Objekte, nach sachlicher Verwandtschaft in Gruppen eingeteilt, die das Ordnungskriterium abgeben. Typische Fragestellungen betreffen die Differenzierung der Menge der Ausdrucksseiten, wobei vielleicht auch Whorfs „sprachliches Relativitätsprinzip“ im Hintergrund eine gewisse Rolle spielt.

Es gibt eine Beziehung zum Thema „Referenz“. Referenz, sprachlich gesehen, ist der Bezug eines Ausdrucks zu einem Gegenstand oder Sachverhalt in der (als gegeben geltenden) Realität. Referenz ist aber nicht mehr in demselben Sinne „statisch“ wie die Ausdrucks-Inhalts-Beziehungen der Wörterbücher. Eine Person referiert unter Verwendung eines sprachlichen Ausdrucks in einem bestimmten Kontext auf einen bestimmten Sachverhalt. Referenz ist dann in dieser Interpretation immer bis zu einem bestimmten Grad individuell bestimmt und an eine jeweilige Situation gebunden. Die Aussagen der Wörterbücher betreffen eine internalisierte Kenntnis (etwa im

Sinne von „Kompetenz“), während Referenz eher auf den Prozess abzielt, der sich dieser Kenntnis bedient (also als zur „Performanz“ gehörig gesehen werden kann). Karl Bühler (1934: 25) hat den Vorgang der Referenz mit dem folgenden – wie er selbst schreibt, damals schon nicht mehr ganz originalen – Beispiel charakterisiert:

„Gesetzt, das Produzieren des Schallphänomens sei im Sprecher angeregt durch einen zeitlich vorausgehenden Sinnesreiz, der von einem Ding im Wahrnehmungsfelde herkommt, und das Hören des sprachlichen Schallphänomens stimuliere den Hörer zur Hinwendung der Augen auf dasselbe Ding. Also zum Beispiel: Zwei Menschen im Zimmer – der eine beachtet ein Prasseln, blickt zum Fenster und sagt: *es regnet* – auch der andere blickt dorthin, sei es direkt vom Hören des Wortes oder sei es vom Blick auf den Sprecher dazu verleitet.“

In ein Schema umgesetzt, hier für die Hörerseite, ergibt das etwa die Anordnung der Abbildung 2.4.1–1.

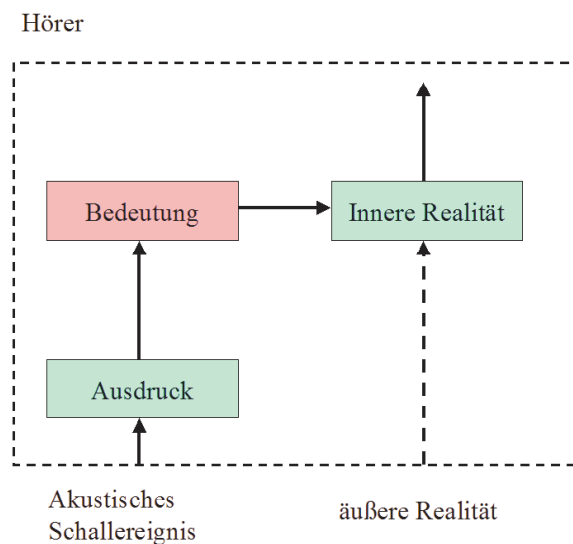


Abbildung 2.4.1–1: Schema zur neuronalen Deutung von „Referenz“.

Wie immer bei Schemata, die Kästen und Pfeile verwenden, muss man sich überlegen, was hier mit diesen Symbolen gemeint ist. Der Inhalt der gestrichelten Umrisslinie muss in unserem Zusammenhang mentale Strukturen und/oder Prozesse repräsentieren. In einer vorsichtigen Interpretation

könnte man das Rechteck „Ausdruck“ als Repräsentation eines Vorgangs sehen, der in der Identifikation des Ausdrucks besteht. Identifikation muss wohl dann heißen: Auswahl aus einer Menge möglicher Ausdrücke. Damit kann „Ausdruck“ auch als Repräsentation eines möglichen Schallereignisses gesehen werden, nicht nur eines Prozesses, denn die Verarbeitungsprinzipien des Kortex lassen eine Trennung von Daten und Prozessen nicht zu. Der Pfeil, der auf die Repräsentation des Ausdrucks zuführt, ist dann nicht mehr nur Andeutung einer prozesshaften Abfolge, sondern erhält eine materielle Interpretation als eine aktivierte Verbindung oder ein Bündel aktivierter Verbindungen zwischen Sinnesorganen und interner Repräsentation. Das kann natürlich nicht heißen, dass auf dieser Verbindung/diesen Verbindungen die dem Schallereignis entsprechende Information in irgend einer Weise kodiert entlangläuft, sondern es ist die *Auswahl* der Verbindung, die diese Information repräsentiert.

Analoges gilt dann auch für die Rechtecke „Bedeutung“ und „innere Realität“ und die darauf zuführenden bzw. von dort ausgehenden Pfeile. Der gestrichelte Pfeil entspricht im Beispiel von Karl Bühler der Wahrnehmung der referenzierten Tatsache durch den Hörer. Die Realisierung der Äußerungsbedeutung ist nicht auf diesen Sinnesinput angewiesen. Die Fortsetzung einer Ereignissequenz kann aber selbstverständlich davon abhängig sein, ob ein entsprechender Sinnesinput erfolgt (durch Kopfbewegung usw. hergestellt werden kann) oder nicht.

Wenn man den Aspekt der Auswahl in die Darstellung mit einbezieht, ergibt sich das Schema der Abbildung 2.4.1–2. In diesem Schema ist auch schon angedeutet, was im Folgenden als Teilbereich dessen, was unter der Überschrift „Semantik“ behandelt werden könnte, Gegenstand der Diskussion sein soll. Negativ formuliert: Nicht die Menge von Verbindungen von den Sinnesorganen zu den Ausdrücken und nicht die Verbindungen von den Ausdrücken zu den Inhalten, sondern der verbleibende Rest, wie immer er auch zu bestimmen sein mag. Wenn man an das Beispiel der Wörterbücher denkt: Es soll gerade nicht die für die Wörterbücher so wichtige Beziehung von Ausdrücken zu Inhalten thematisiert werden. Dieser Gegenstand wird in Kapitel 4 („Lexikon“) eingehender behandelt. Sondern es soll die Struktur und Funktion dessen abgeklärt werden, was die Inhalte selbst ausmacht. Jeder Linguist wird die Vorstellung akzeptieren, dass Inhalte nicht unabhängig voneinander spezifiziert zu denken sind, so dass man die Darstellung durch Bündel separater Rechtecke in Abbildung 2.4.1–2 sicherlich nicht „wörtlich“ nehmen kann. Das ausführlich zu diskutierende Hauptproblem ist aber das Verhältnis von Bedeutung und innerer Realität oder, mit

anderen Vokabeln und mit Bezug auf eher statisch gedachte Informationsbestände ausgedrückt: zwischen Semantik und Weltwissen.

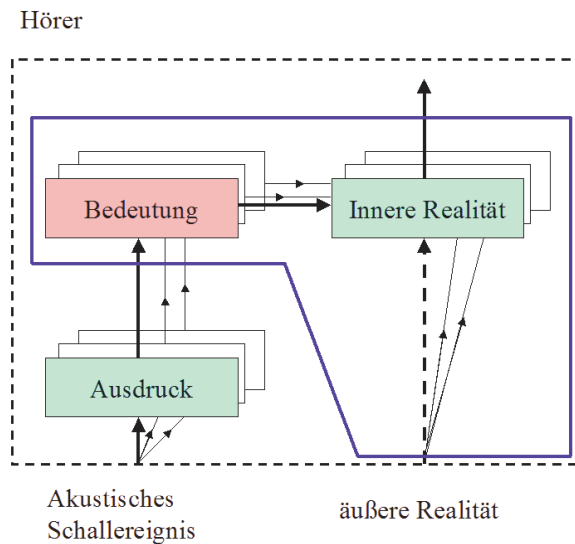


Abbildung 2.4.1–2: Erweitertes Schema zur neuronalen Deutung von „Referenz“. Die fette blaue Linie deutet den Gegenstandsbereich dieses Kapitels an.

Von dem Startpunkt der Wörterbücher ausgehend, ist hier bisher vor allem der lexikalische Bereich angesprochen und beachtet worden, das heißt ein Bereich, in dem es um elementare bedeutungshaltige Ausdrücke geht. Ein komplexer Satz ist nicht elementar in diesem Sinne. Elementare bedeutungshaltige Ausdrücke sind gleichzeitig solche, für die durch Lernprozesse entstandene Repräsentationen im Kortex bereitgehalten werden. Ganze Sätze können zu elementaren bedeutungshaltigen Ausdrücken werden, wenn sie als Ganzes im Kortex durch Lernprozesse verankert sind. Das Schema der Abbildung 2.4.1–2 deckt nur solche elementaren Ausdrücke und ihre Verarbeitung ab. Es ist damit auch ausgesagt, dass die folgende Argumentation zunächst auf diesen Bereich beschränkt bleibt. Die Perspektive ist damit die der „lexikalischen Semantik“. Das Problem der Bedeutung und des Verstehens von Sätzen und Texten wird in Teil 6, „Gedächtnisformen, Textverstehen“ behandelt.

2.4.2 Bedeutungsbeschreibung und Realitätsbezug

In einer Veröffentlichung, die *vor* der Rezeption prototypensemantischer Ideen liegt, schreibt John Lyons:

„Es ist wahrscheinlich richtig zu sagen, daß die Mehrheit der strukturellen Semantiker heutzutage die eine oder andere Version der Komponentenanalyse vertritt. Dieser Ansatz zur Beschreibung der Bedeutung von Wörtern beruht auf der These, daß die Bedeutung eines jeden Lexems aufgrund einer Menge allgemeinerer Bedeutungskomponenten (oder semantischen Merkmale) analysiert werden kann, von denen einige oder alle mehreren verschiedenen Lexemen im Wortschatz angehören.“ (Lyons, 1980: 327)

In dem von Lyons an derselben Stelle gegebenen Beispiel würde also der Inhalt des Wortes „Frau“ als Kombination der Komponenten „weiblich“, „erwachsen“ und „menschlich“ zu beschreiben sein.

Es ist vielleicht nicht ohne weiteres legitim, solche aus dem strukturalistischen Paradigma stammenden Aussagen auf kortikale Strukturen zu beziehen, das ist aber notwendig, wenn wir in unserem Zusammenhang Gebrauch davon machen wollen.

Aber auch losgelöst von dieser Geltungsproblematik stellt sich nicht nur die Frage, mit welcher Begründung bestimmte Komponenten bei der Analyse eines einzelnen Begriffs diesem Begriff zugeschrieben werden, sondern noch viel elementarer, woher solche Komponenten überhaupt bezogen werden können. Um der Kritik zu begegnen, dass in den entsprechenden Bedeutungsbeschreibungen doch nur Wörter einer bestimmten Sprache durch andere Wörter ggf. derselben Sprache erklärt würden und man sich damit also im Kreis bewege, ist auf den Unterschied von „Objektsprache“ und „Metasprache“ hingewiesen worden, was wahrscheinlich wenig sinnvoll ist, wenn man nicht zusätzlich die Möglichkeit einer universellen Beschreibungssprache als Metasprache annimmt (vgl. Katz & Fodor, 1963: 208), begründet etwa nach den Vorstellungen bei Lyons (1980: 342):

„Wenn einige (aber keineswegs alle) der semantischen Unterscheidungen, die in Sprachen gemacht werden, durch eine genetisch übermittelte Disposition, auf biologisch und kulturell auffallende Stimuli zu reagieren, determiniert sind, dann werden Sprachen allmählich diese semantischen Unterscheidungen lexikalisieren (und vielleicht auch grammatikalisieren): z. B. den Unterschied zwischen dem, was senkrecht ist, und dem, was nicht senkrecht ist, zwischen dem, was

fest ist und dem, was nicht fest ist, zwischen dem was belebt ist und dem, was nicht belebt ist.“

In engem Zusammenhang mit dem Problem der Universalität steht das Problem der Atomizität der semantischen Komponenten. Wenn Bedeutungen durch Komponenten beschrieben (definiert) werden, ist implizit vorausgesetzt, dass die Komponenten nicht ihrerseits einer analogen Behandlung bedürfen, mindestens aber, dass der Vorgang des fortgesetzten Definierens mit dem Erreichen atomarer und nicht zu definierender Komponenten abbricht.

Der Standpunkt der Atomizität und Universalität semantischer Primitive ist am entschiedensten von Anna Wierzbicka vertreten worden.

„The elements which can be used to define the meaning of words (or any other meanings) cannot be defined themselves; rather, they must be accepted as “indefinibilia”, that is, as semantic primes, in terms of which all complex meanings can be coherently represented.“
Wierzbicka (1996: 10)

Nicht-definierbare „Primes“ dieser Art müssen angeboren sein, und sie sind deshalb universell. Wierzbicka (1996: 13f.):

„In the theory presented in this book it was hypothesized, from the start, that conceptual primitives can be found through in-depth analysis of any natural language; but also, that the sets of primitives identified in this way would “match”, and that in fact each such set is just one language-specific manifestation of a universal set of fundamental human concepts.

For example, it was expected that the concepts ‘someone’, ‘something’, and ‘want’, which are indefinable in English, would also prove to be indefinable in other languages; and that other languages, too, will have words (or bound morphemes) to express these concepts.

This expectation was based on the assumption that fundamental human concepts are innate, in other words that they are part of the human genetic endowment; and that if they are innate, then there is no reason to expect that they should differ from one human group to another.“

Die Primes zusammen mit einer Syntax bilden bei Wierzbicka die „Natural Semantic Metalanguage“. Bedeutungen sind vollständig in dieser Sprache zu

beschreiben. (Wortbedeutungen bestehen nicht in einer bloßen Aufzählung von Primes, sondern sind durch komplexere Strukturen wiederzugeben.)

Zu den Primes gehören GOOD und BAD. Sie sind deshalb interessant, weil es Hinweise darauf gibt, dass tatsächlich bestimmte Hirnbereiche (z. B. die Amygdala) für dergleichen Bewertungen zuständig sein könnten, sodass man hier der Angeborenheitsthese noch am ehesten zustimmen würde.

Was ergibt sich daraus z. B. für die Repräsentation der dem Wort „gefährlich“ entsprechenden Bedeutung im Kortex? BAD muss eine Komponente der Bedeutung von „gefährlich“ sein. Wenn das Wort „gefährlich“ wahrgenommen wird, müssen ggf. Handlungen im Hörer ausgelöst werden, um erwarteten Schaden zu vermeiden. BAD darf – jedenfalls im eigentlichen Gebrauch – nicht etwas prinzipiell Fiktives sein. Damit BAD eine Wirkung haben kann, muss es eine Wahrnehmung geben, die BAD entspricht. Wenn BAD durch Großmutterzellen repräsentiert gedacht wird, muss es Verbindungen von Wahrnehmungsorganen zu diesen Großmutterzellen geben. Dafür kommen z. B. Schmerzsensoren, aber auch Sensoren für andere (negative) Wahrnehmungen in Frage. Damit ist BAD durch bestimmte Verbindungen mit der Sinnesperipherie definiert und kann zur Bildung komplexerer Konzepte genutzt werden.

Wie steht es aber mit BIG oder KIND oder ALIVE? Man beachte, dass BIG ein relatives Konzept ist (Wierzbicka, 1996:54) und KIND („a rose is ‘a kind of flower’“; Wierzbicka, 1996:62) eine abstrakte Begriffsbeziehung ausdrückt. Kann man sich Wahrnehmungsorgane vorstellen, die solche Primes definieren?

Die Frage kann offen bleiben. Wichtig ist in unserem Zusammenhang, dass Primes nur insofern akzeptiert werden können, als periphere Organe der Wahrnehmung oder der Motorik eine Art von Definition leisten. Es kann keine angeborene Großmutterzelle geben, die ohne entsprechende Verbindungen mit der Peripherie weiß, für welche Wahrnehmungen oder Aktionen der Peripherie sie steht. Dasselbe gilt für Zellkomplexe. Isolierte Zellkomplexe können keine Bedeutung tragen.

In diesem Zusammenhang ist es interessant, das Beispiel der „semantischen Netze“ heranzuziehen, am besten in der ursprünglichen Form bei Quillian (1968). Quillian definiert Begriffe in netzförmigen Strukturen, die „tokens“ anderer Begriffe enthalten.

Das führt zu Strukturen wie der folgenden aus Quillian (1968:225), um den Vergleich zu erleichtern aus der dort gegebenen graphischen Darstellung in einen definierenden Text umgesetzt und vereinfacht:

plant
 Is a structure modified by live and not animal and frequently with
leaves and getting food from air or water or earth

Das Oval kennzeichnet einen „type node“, die Rechtecke kennzeichnen „token nodes“. Token nodes können verstanden werden als Zeiger auf (bzw. Verbindungen mit) type nodes und die (den) damit verbundenen Informationen, die wieder token nodes enthalten. Die nicht als token nodes ausgezeichneten Elemente sind wohl als universell und nicht einer Definition bedürftig anzusehen („not“ und „frequently“ werden auf einer Intensitätskala als 0 bzw. 7 kodiert).

Der wesentliche Unterschied zu den Definitionen bei Wierzbicka ist nicht, dass überhaupt komplexe Begriffe zur Definition verwendet werden, die ihrerseits an anderer Stelle definiert sind. Hierarchische Definitionen werden von Wierzbicka akzeptiert (vgl. Wierzbicka, 1996: 211 ff.). Ein Unterschied besteht aber in der Zahl und der Art der Primitive.

Ein entsprechendes Beispiel aus Wierzbicka (1996: 218) (Oval und Rechteck sind zum besseren Vergleich mit Quillian hinzugefügt):

face
 „ a part of a person's head
 it is on one side of the head
 it has parts
 often, when a person feels something,
 something happens in this part
 other people can see this
 because of this, when a person feels something,
 other people can often know something about it“

In dieser Definition ist „head“ ein nicht-primitives Konzept, das eine eigene Definition erhält, die übrigen Elemente können als primitiv angesehen werden. Es gibt durchaus primitive Konzepte, die Wörtern wie „person“, „people“, „something“ usw. entsprechen.

Man vergleiche damit Quillian(1968:228): „Moreover, there are no word concepts as such that are “primitive.”“ Das bedeutet, dass alle verwendeten „Wort-Konzepte“ bei Quillian immer ihrerseits durch wieder andere Konzepte definiert sind, so dass man bei der Verfolgung eines Pfads solcher

Definitionen nicht irgendwann auf nicht mehr definierte Primitive stößt. Bei Wierzbicka werden von vornherein auch bestimmte Konzepte, die Wörtern entsprechen, als primitiv, also atomar betrachtet, so dass der Pfad einer Definition an diesen Stellen abbricht.

Allerdings findet sich bei Quillian im Rahmen einer Diskussion über zusätzliche Erfordernisse für das Erkennen von Objekten die folgende Bemerkung:

„But, a network containing one-way associative links from an object's name to the set of properties of that object (as ours does now) would seem already to contain all the *nodes* needed to recognize a particular object given its sensed properties.“ (Quillian, 1968: 229; Unterstreichung hinzugefügt)

Hier werden quasi nebenbei Sinnesdaten als Voraussetzung für ein Funktionieren des Netzes eingeführt. Es ist auch klar, dass solche Sinnesdaten für die Bedeutungsdefinition Primitive sind. Ein semantisches Netz kann, wenn man an das Erkennen (Kategorisieren) von Objekten denkt, nur dann funktionieren, wenn es Verbindungen zur „Außenwelt“ aufweist. Ohne solche Verbindungen bleiben die Definitionen zirkulär. Das Netz ist in sich geschlossen.

Man kann nun, wie das Jean Aitchison (1987/2003:70) tut, diese Zirkularität akzeptieren und die Ansicht vertreten, dass es Wunschdenken ist, dass die Reise durch die Definitionen an bestimmten Punkten (die dann elementar sind) endet. Aber es ist nicht nur die formale Eigenschaft der Zirkularität zu beachten. Ein semantisches Netz ist, sofern die Verbindung mit der Sinneswahrnehmung nicht hergestellt wird, metaphorisch gesprochen „blind“. Dasselbe gilt auch von den Definitionen bei Wierzbicka, solange die Primitive nicht als angeborene Kategorien der Sinneswahrnehmung gelten können. Dass es angeborene Sinnesprimitive gibt, kann angesichts der beobachtbaren biologischen Strukturen und der lokalistischen Natur der peripheren Repräsentation nicht bezweifelt werden. (Das Problem abstrakter Begriffe, die ja nicht in direkter Weise auf Sinneswahrnehmungen zurückgeführt werden können, wird unten in Kapitel 2.5 „Vorstellungen“ behandelt. Man beachte auch, dass nicht von Wortformen, also Ausdrucksseiten von Wörtern die Rede ist. Es ist prinzipiell denkbar, dass es Ausdrucksseiten gibt, denen keine Bedeutung zukommt. Die Ausdrucksseiten selbst sind über Verbindungen zur auditiven Wahrnehmung und zur Motorik definiert.)

Damit bewegen wir uns mit der gesamten Problematik der Fundierung semantischer Beschreibungen in einem Bereich, der mehr oder weniger direkt in Verbindung mit der Sinneswahrnehmung steht, also in einem Bereich, der,

in die Abbildung 2.4.1-2 eingetragen, eine Eingrenzung wie in Abbildung 2.4.2-1 ergibt.

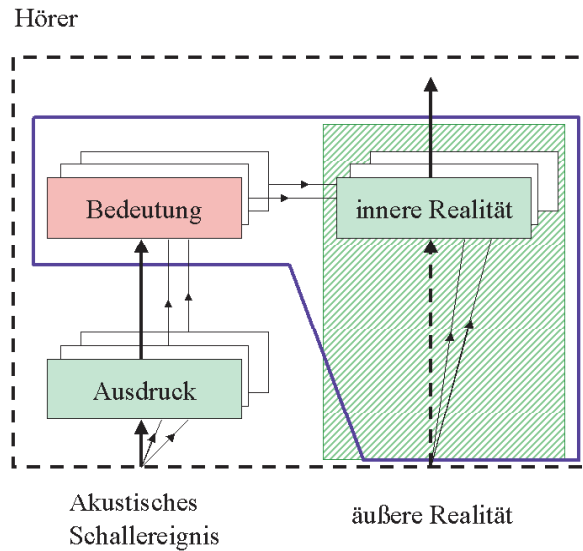


Abbildung 2.4.2-1: Erweitertes Schema zur neuronalen Deutung von „Referenz“. Das schraffierte Rechteck bezeichnet den Bereich, der von Fragen nach semantischen Primitiven mit Sicherheit betroffen ist.

Das in der Abbildung mit „Bedeutung“ beschriftete Rechteck kann nur insofern an den Phänomenen beteiligt sein, als es in Verbindung mit dem Bereich der „inneren Realität“ steht.

2.4.3 Kategorisierung

Nach den Schlussfolgerungen des vorigen Abschnitts liegt es nahe, sich etwas genauer mit dem Vorgang der Kategorisierung zu beschäftigen. Kategorisierung meint den Übergang von Sinnesdaten oder mindestens indirekt damit verknüpfbaren Informationen zu internen Repräsentationen. Diese internen Repräsentationen entsprechen Kategorien insofern, als sie gegenüber der Menge der Sinnesdaten ein gewisses Maß an Abstraktion einschließen, sodass verschiedene Bündel von Sinnesdaten einer und derselben internen Repräsentation zugewiesen werden.

Der damit angesprochene Zuweisungsprozess ist in neuronaler Interpretation im einfachsten Fall identisch mit einem Wahrnehmungsprozess. Das Ergebnis eines Wahrnehmungsprozesses ist nicht damit zufriedenstellend beschrieben, dass man sagt, der Wahrnehmungsprozess führe zu einem spezifischen Zustand des Gehirns. Man muss auch erklären, warum dieser Zustand nicht einfach identisch ist mit einer direkten Spiegelung der Aktivität der Sinnesorgane, also, etwas anders ausgedrückt, man muss erklären, welche Funktion die internen, an die Sinneswahrnehmung im engeren Sinne anschließenden Verarbeitungsprozesse haben.

Die direkte Spiegelung der Aktivität der Sinnesorgane wäre so etwas wie eine extreme Form verteilter Repräsentation einer Kategorie, die eben in der Gesamtheit dieser Aktivität besteht. Man stelle sich ein künstliches neuronales Netz vor, in dem die Einheiten der Eingabeschicht durch jeweils eine einzige Verbindung mit entsprechenden Einheiten der Ausgabeschicht verbunden sind. Die Repräsentation der Eingabeschicht wird lokalistisch, die der Ausgabeschicht – per Definition – verteilt interpretiert: eine sichtlich sinnlose Vorstellung.

Es ist nicht nur so, dass ein beobachtender Wissenschaftler zu unterscheiden hat darüber, ob zwei Zustände des Gehirns gleich oder verschieden sind, sondern es muss innerhalb des Gehirns selbst die Möglichkeit geben, dass verschiedene Repräsentationen des sensorischen Inputs zu verschiedenen Reaktionen auf der Seite des motorischen Outputs führen können. Man beachte, dass die Gesamtheit des Sinnesinputs von Millisekunde zu Millisekunde wechselt. Auch dieser andauernde Wechsel, nicht nur die Tatsache, dass die Wiederholung des exakt Gleichen sehr unwahrscheinlich ist, macht einen Teil des Kategorisierungsproblems aus.

Aus der in Kapitel 2.3 verteidigten Annahme lokalistischer Repräsentation im Gehirn ergeben sich die geforderten Leistungen als natürliche Konsequenz. Eine Kategorie ist eine Einheit, redundant in Großmutterzellen repräsentiert. Es ist nicht so, dass ein Konzept (eine Kategorie) in einer Ansammlung von Merkmalen besteht, sondern es ist zusätzlich(!) ein zusammenfassendes Neuron oder – realistischer – es sind mehrere bis viele gleichwertige zusammenfassende Neuronen vorhanden. Aus der Aktivität der zusammenfassenden Neuronen können problemlos entsprechende Reaktionen abgeleitet werden.

Die Annahme, dass die Kategorie-als-Einheit (gemeint ist eine Kategorie *oberhalb* der Sinnesprimitive) atomar und damit angeboren ist, ist deutlich unhaltbar. Es ist unklar, ob selbst einer der Hauptvertreter solcher Thesen, Jerry A. Fodor, heute diesen Standpunkt noch wirklich vertritt. Fodor (1998: 162):

„Likewise, it used to seem to me that atomism about concepts means that DOORKNOB is innate. But now I think that you can trade a certain amount of innateness for a certain amount of mind-dependence. *Being a doorknob* is just: striking our kinds of minds the way that doorknobs do. So, what you need to acquire the concept DOORKNOB “from experience” is just: the kind of mind that experience causes to be struck that way by doorknobs.“

Wenn der Vorgang der Kategorisierung voraussetzt, dass entsprechende Kategorien bzw. Konzepte zuvor gelernt worden sind, gelten auch die in Abschnitt 2.3.4 formulierten Bedingungen für Lernprozesse. Lernprozesse liefern aus den dort dargestellten Gründen hierarchisch gegliederte Strukturen, das heißt, dass ein Konzept aus primitiveren Konzepten und diese wieder aus noch primitiveren Konzepten aufgebaut sein kann.

Die damit erreichte Parallelität der Strukturen mit dem, was in Abschnitt 2.4.2 über die Struktur von Wortbedeutungen gesagt ist, ist augenfällig. Kategorien sind konzeptuelle Einheiten, bestehend aus Hierarchien von Komponenten und fundiert in angeborenen Primitiven. Es ist völlig offen, wie eine lexikalische Bedeutung spezifiziert sein könnte ohne direkten Bezug auf diese Strukturen. In Kochendörfer (2002) ist darauf hingewiesen worden, dass schon die Frage der Lernbarkeit lexikalischer Bedeutungen und die Tatsache, dass die Kategorisierung im Spracherwerb dem Erwerb lexikalischer Ausdrucksseiten vorausgeht, die Schlussfolgerung erzwingen, dass eine von der Kategorisierung abgekoppelte Repräsentation lexikalischer Bedeutungen nicht erklärbar wäre.

Die bei Aitchison (1987/2003:41) relativ vorsichtig formulierte These, dass Wortbedeutungen u. U. nicht vollständige Konzepte darstellen, ist oberflächlich gesehen damit nicht ohne weiteres vereinbar:

„Second, we will treat the ‘meaning’ of a word as overlapping with the concept to a large extent, though not necessarily totally: the overall concept may extend beyond the sections labelled with a word (figure 4.2).“

Diese Schwierigkeit verschwindet aber, wenn man beachtet, dass selbstverständlich nicht alle Konzepte tatsächlich sprachliche Ausdrucksseiten haben, sodass also gegenüber sprachlich realisierten Konzepten auch in gewissem Sinn „vollständigere“ Konzepte möglich sind.

Wir kommen insgesamt zu der – vielleicht im ersten Augenblick unbequem – Folgerung, dass eine gegenüber der außersprachlichen Kategorisierung selbständige linguistische Semantik keine neuronale Realität haben kann.

Wie in Abbildung 2.4.3–1 dargestellt, entfällt damit die Wortbedeutung als eigenständige Instanz.

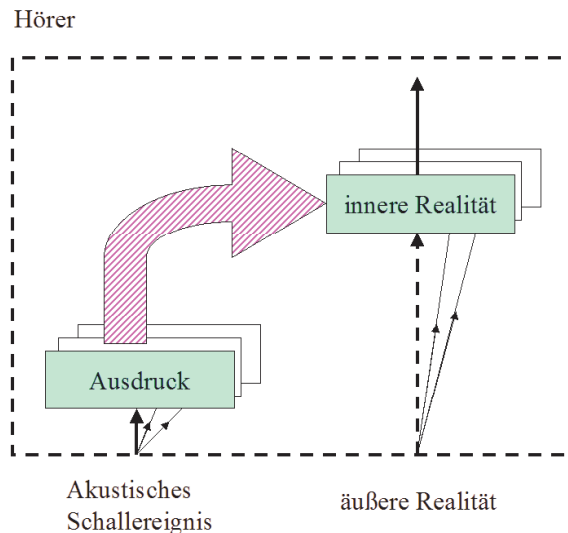


Abbildung 2.4.3–1: Schema zur neuronalen Deutung von „Referenz“ ohne eigenständige linguistische Semantik. Sprachliche Ausdrucksseiten müssen, wie durch den Blockpfeil angedeutet, direkt mit Wahrnehmungskonzepten verknüpft gedacht werden.

2.4.4 Wortfelder, Bedeutungsverwandtschaft und distinktive Merkmale

Man kann leicht Auskünfte von Versuchspersonen über Bedeutungsverwandtschaften erhalten. Wenn man nach nächstverwandten Konzepten bzw. Wörtern fragt, kann man z. B. für „Kanne“ die Auskunft „Krug“ bekommen. Solche Auskünfte sind auch zwischen verschiedenen Personen – vielleicht nach einigem Hin- und Her – konsensfähig. Schwieriger wird es, wenn man versucht, die Eigenschaften (die semantischen Merkmale) zu ermitteln, die, in unserem Beispiel, das Konzept KANNE von dem Konzept KRUG unterscheiden. Dafür wird man u. U. verschiedene Auskünfte bekommen. Allerdings handelt es sich bei den dabei genannten Merkmalen typischerweise um solche, die als Unterscheidungskriterium für viele Kannen und Krüge gleich gut geeignet, also in einem gewissen Sinn austauschbar sind.

Die Frage nach dem unterscheidenden Merkmal ist wohl nur sinnvoll, wenn man annimmt, dass es prinzipiell auch Merkmale geben kann, die für die erfragten Konzepte *gemeinsam* gelten. In der strukturalistischen Semantik werden auch diese Merkmale durch die Feststellung von Bedeutungsunterschieden gewonnen, sie sind in diesem Sinne distinktiv. Einigkeit besteht, auch bei Informanten, darüber, dass bestimmte faktisch vorhandene Merkmale (z. B. das Blümchenmuster auf der Kanne) für das Konzept irrelevant sind.

Strukturen bedeutungsverwandter Wörter heißen „Wortfelder“. Da es sich eigentlich um Strukturen von Inhaltsseiten, nicht Ausdrucksseiten von Wörtern handelt, bieten sich auch andere Bezeichnungen an, es wird z. B. von „Begriffsfeldern“ gesprochen. Einige wesentliche Komponenten der Wortfeldtheorie, die sich allesamt aus Grundpositionen der strukturalistischen Sprachtheorie ergeben, werden bei dem Klassiker Jost Trier (1931/1973) so charakterisiert:

„Das Wortfeld ist zeichenhaft zugeordnet einem mehr oder weniger geschlossenen Begriffskomplex, dessen innere Aufteilung sich im gegliederten Gefüge des Zeichenfeldes darstellt, in ihm für die Angehörigen einer Sprachgemeinschaft gegeben ist. Dies in einem inhaltlich zusammengehörigen Teilausschnitt des Wortschatzes — eben dem Wortfeld — sichtbar werdende Gefüge ist also die äußere, zeichenhafte Seite der begrifflichen Aufteilung, die jener Begriffskomplex, Begriffsbereich, Begriffsbezirk durch und für die Sprachgemeinschaft erfährt. Die das Wortfeld, den Wortmantel, die Wortdecke mosaikartig zusammensetzenden Einzelwörter legen — im Sinne ihrer Zahl und Lagerung — Grenzen in den Begriffsbereich hinein und teilen ihn auf.“ (S. 1)

„Dies Begriffsfeld gewinnt keine körperlich wahrnehmbare Erscheinungsform in äußerer Lautung, deswegen ist es aber nicht weniger wirksam, nicht weniger einwirkend auf den Sinn des Wortes, nicht weniger dem Bewußtsein zugänglich. Oberhalb und außerhalb des vom Einzelmenschen gesprochenen Satzes steht eben die große Wirklichkeit des sprachlichen Zeichensystems mit seinen vielen Begriffsfeldern, unter deren sinngemäßer Benutzung jeder einzelne zu hör- und sichtbarer Gestalt gelangende Satz allererst aufgebaut werden muß. Wir haben das Wort nur im Zusammenhang dieses Ganzen. Das Wortzeichenfeld als Ganzes muß gegenwärtig sein, wenn das einzelne Wortzeichen verstanden werden soll, und es wird verstanden im Maße der Gegenwärtigkeit des Feldes. Es „bedeutet“ nur in diesem Ganzen und kraft dieses Ganzen. Außerhalb eines Feldganzen kann es ein Bedeuten überhaupt nicht geben.“ (S. 4)

Die Vorstellung des mosaikartigen, lückenlosen Wortmantels bei Trier hat ihr Äquivalent in der Idee, dass Bedeutungen durch distinktive Merkmale definierbar sind. Aus der Lückenlosigkeit ergibt sich, dass Veränderungen der „Ausdehnung“ eines Konzepts Veränderungen der Nachbarkonzepte mit sich bringen. Gleichzeitig wird die Aussage verständlich, dass die Bedeutung eines Worts von der Gesamtheit der Wortbedeutungen abhängig ist. Schwierig ist die Behauptung (auch außerhalb der zitierten Textstelle), dass die Gesamtheit der Bedeutungen anklingt, wenn eines der Konzepte aktiviert wird. Man muss so interpretieren, dass, ausgehend von einem bestimmten Konzept, Vorstellungen entwickelt werden können, die den verwandten Konzepten entsprechen. Auf die an sich interessanten Querbeziehungen zu Traditionen des „semantischen Holismus“ in der Philosophie soll hier nicht eingegangen werden (vgl. Esfeld, 2002 und Mayer, 1997).

Schließlich ist es nur noch ein Schritt von den Vorstellungen Triers zu dem „sprachlichen Relativitätsprinzip“ Whorfs:

„Wir gelangen daher zu einem neuen Relativitätsprinzip, das besagt, daß nicht alle Beobachter durch die gleichen physikalischen Sachverhalte zu einem gleichen Weltbild geführt werden, es sei denn, ihre linguistischen Hintergründe sind ähnlich oder können in irgendeiner Weise auf einen gemeinsamen Nenner gebracht werden (be calibrated).“ (Whorf, 1963/1999:12)

Die Wahrnehmung der Welt ist von den sprachlich verfügbaren Kategorien abhängig. (Zu Whorf und zu anderen Autoren, die mit dem Komplex der sprachlichen Relativität verknüpft werden können, vgl. Werlen, 2002.)

Der damit in einigen wesentlichen Punkten skizzierte Gesamtzusammenhang der lexikalischen Semantik in strukturalistischer Sicht enthält als Kern die Idee der Bedeutungsverwandtschaft und der Bedeutungsdifferenz (Distinktivität, z. B. auf der Merkmalsebene). Bedeutungsverwandtschaft ist, wenn man an neuronale Konzeptbildung denkt, die die Grundlage von Kategorisierungsvorgängen ist, problemlos und selbstverständlich. Sie entsteht zwangsläufig schon dadurch, dass verschiedene Konzepte der Wahrnehmung eine gewisse Menge von Sinnesprimitiven gemeinsam haben können (und in großem Umfang auch gemeinsam haben müssen). Auch gemeinsame nicht mehr primitive Komponenten sind in dem Augenblick selbstverständlich und neuronal plausibel, wo überhaupt mit einer hierarchischen Konzeptbildung gerechnet wird. Komponenten, die ein Konzept bilden, können auch auf verschiedenen Verarbeitungsebenen angesiedelt sein wie in Abbildung 2.4.4–1 angedeutet. Eine mögliche Beschränkung für solche Mischformen ergibt sich aus den Laufzeiten der Aktionspotenziale, da die Aktionspo-

tenziale ja in einem relativ schmalen Zeitfenster bei den zu definierenden Großmutterzellen wirksam werden müssen. Es versteht sich von selbst, dass es neuronal gesehen (über neuronale Verbindungen realisiert) keine Bedeutungsverwandtschaft semantischer Primitive geben kann (gegen Wierzbicka 1996). Die Repräsentation von Bedeutungsverwandtschaft, auch oberhalb der Ebene der Primitive, durch direkte Verbindungen von Großmutterzellen der verwandten Konzepte ist in einem System, in dem Einzelimpulskodierung gilt, schon aus Gründen der Lernbarkeit nicht denkbar.

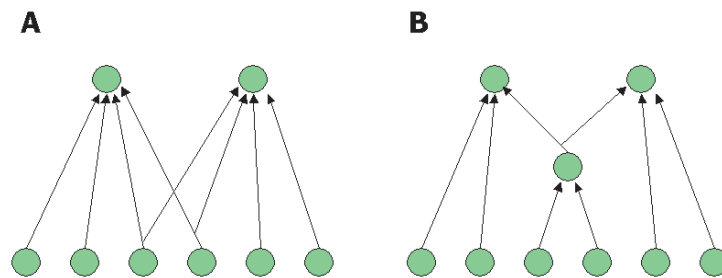


Abbildung 2.4.4–1: Bedeutungsverwandtschaft durch gemeinsame Komponenten. **A.** Alle Komponenten sind elementar. **B.** Mischform mit elementaren und komplexen Komponenten.

Viel schwieriger als die Bedeutungsverwandtschaft ist die Distinktivität zu verstehen. Sie setzt die Unterscheidung von relevanten und irrelevanten oder bestenfalls redundanten Merkmalen voraus. Ein irrelevantes Merkmal kann ein Merkmal sein, das über die Sinneswahrnehmung erfasst wird, das aber nicht zur Definition einer der Wortbedeutung entsprechenden Großmutterzelle verwendet wird. Wenn redundante Merkmale eine Funktion zur Sicherung der Kategorisierung haben sollen, müssen sie alternativ zu anderen Merkmalen Geltung haben.

Wenn man die Unterscheidung auf diese Weise konstruiert, wird deutlich, dass ein zusätzliches Argument gebraucht wird: Es ist natürlich, wenn man nur eine einzelne Person betrachtet, die aufgrund von Sinneserfahrungen Konzepte bildet, beliebig und von vielen Umständen abhängig, welche Merkmale zur Kategorisierung herangezogen werden. Die Idee der strukturalistischen Semantik ist aber, dass die Konzepte innerhalb einer Personengruppe oder letztlich einer Sprachgemeinschaft sich so weit ähneln, dass eine Kommunikation möglich wird. Relevante Merkmale sind dann solche, die den Konsens ausmachen. Merkmale, die den Konsens ausmachen, müssen vorhanden sein und ausreichen, um ein benanntes Exemplar einer entsprechenden Kategorie zuzuordnen.

Das Etablieren einer sprachlichen Verständigungsbasis ist, ganz allgemein gesprochen, die Aufgabe des Spracherwerbsprozesses. Man kann beobachten, dass Kinder im Spracherwerbsprozess Wörter überspezifiziert verwenden, das heißt, nur für bestimmte Exemplare der Kategorie, die in der Erwachsenensprache mit dem lexikalischen Ausdruck verbunden ist. Es sind, bezogen auf die Erwachsenensprache, „überschüssige“ Merkmale da. Die Anpassung an die Erwachsenensprache kann dann dadurch geschehen, dass weitere Konzepte gebildet werden, darunter können dann weitere überspezifizierte sein, so dass das Problem durch eine Art Mehrdeutigkeit der Ausdrucksseite gelöst wird.

Interessanter ist das entgegengesetzte Problem, die Unterspezifizierung von Kategorien, die für die jeweiligen Bezugspersonen auch viel auffälliger ist. Um hier ein charakteristisches Beispiel zu geben: Ein Kind zeigt an der Straßenbahnhaltestelle auf die heranfahrende Straßenbahn und ruft „Zug, Zug, Zug“. Die begleitende Mutter korrigiert: „Das ist kein Zug, das ist eine Straßenbahn“. Wir lassen außer Acht, was das für den Erwerb der Ausdrucksseite „Straßenbahn“ durch das Kind bedeutet. Offenbar verfügt das Kind über ein Konzept ZUG, das auf das Erscheinungsbild einer Straßenbahn passt, und man kann annehmen, dass dieses Konzept durch Großmutterzellen redundant repräsentiert ist.

Es ist nun nicht möglich, eine Anpassung an die Erwachsenensprache dadurch zu erreichen, dass eine oder einige dieser Großmutterzellen so verändert werden, dass sie eine zusätzliche differenzierende Verbindung bekommen und die anderen Verbindungen in ihren Gewichten entsprechend angepasst werden. Dergleichen Vorgänge würden einen übergeordneten neuronalen Mechanismus voraussetzen, der in der Lage ist, neuronale Verbindungen nach nicht-lokalen Kriterien zu verändern. Also bleibt nur, dass ein neues Konzept, repräsentiert durch eine bisher nicht definierte Großmutterzelle, ggf. unter Verwendung des Konzepts ZUG gebildet wird, z. B. ZUG + FÄHRT IN DER STADT. Es ist nicht gewährleistet, dass ein ganz bestimmtes ergänzendes Merkmal verankert wird, wichtig ist nur, dass es zur Differenzierung von ZUG und STRASSENBAHN taugt.

Solche Korrekturprozesse sind an eine Auffälligkeit des „Fehlers“ gebunden. Sie gewährleisten, dass die Unterscheidungen, die eine Sprachgemeinschaft bezüglich der Realität der Sinneswahrnehmung macht, von den Mitgliedern der Sprachgemeinschaft in ihre Kategorisierung aufgenommen werden. In diesem Sinne stoßen die Bedeutungen im Sprachsystem bildlich gesprochen „räumlich“ aneinander: Die Sprachgemeinschaft ist bemüht, bestimmte Bedeutungen auseinanderzuhalten und sie tut es in der beschriebenen Weise. Man kann so Triers Mosaikmetapher verstehen, ohne gezwungen zu sein,

die These von der Lückenlosigkeit des Wortfelds zu akzeptieren, da die Bedeutungen (die Kategorien) auch dann bestehen, wenn sie nicht gegenüber Nachbarn abgegrenzt werden können.

In Kochendörfer (2002) wird diese Überlegung durch das Argument ergänzt, dass Kategorien, die sprachlich realisiert werden, durch die sprachliche Benutzung eine höhere Gebrauchsfrequenz aufweisen und damit weniger dem Vergessensprozess unterliegen als die sprachlosen. Auf diese Weise kann das Kategoriensystem einer Person auf sprachlich realisierte Kategorien hin korrigiert werden. Wenn aber Spracherwerbsprozesse auf die Kategorisierung Einfluss nehmen und die Semantik eben nicht eine rein sprachliche Angelegenheit ist, wird selbstverständlich, dass wir mit einer sprachlichen Relativität nach den Vorstellungen Whorfs, mindestens im lexikalischen Bereich, tatsächlich rechnen müssen.

2.4.5 Prototypizität

Der Kern der Prototypentheorie in der Semantik besteht in Experimenten, in denen Versuchspersonen dazu gebracht werden, bestimmte Begriffe danach zu beurteilen, wie gut sie in eine vorgegebene Kategorie passen. Ein klassisches Beispiel: Vorgegeben wird die Kategorie „Obst“. Die Versuchspersonen werden aufgefordert, für Begriffe wie „Apfel“, „Heidelbeere“, „Tomate“ usw. auf einer Skala von 1 bis 7 anzugeben, wie gut sie in die Kategorie „Obst“ passen. Das Ergebnis solcher Versuche ist, dass es offenbar „bessere“ und „schlechtere“ Exemplare mindestens für viele, wenn nicht prinzipiell für alle Kategorien gibt. Die Einschätzungen sind auch von Person zu Person relativ stabil. (Zu entsprechenden Experimenten vgl. z. B. Rosch, 1975.) Man kann die Menge der Exemplare als eine Art „fuzzy set“ verstehen: Es ist nicht so, dass ein Exemplar zu einer Kategorie gehört (den Zugehörigkeitswert 1 hat) oder nicht (den Zugehörigkeitswert 0 hat), sondern die Zugehörigkeit kann beliebige Werte zwischen 0 und 1 annehmen. (Ein Versuch, die Fuzzy-Set-Theorie direkt zur Bewältigung des Problems semantischer „Unschärfe“ einzusetzen, wird von Burghard Rieger in Rieger (1989) unternommen. Eine Umsetzung in eine neuronale Architektur ist aber nicht beabsichtigt und sicherlich auch nicht möglich.) Das am besten bewertete Exemplar, für die Kategorie „Obst“ in Mitteleuropa der Apfel, wird als Prototyp bezeichnet. Soweit ein solcher Prototyp existiert, kann man die Bedeutung eines Exemplars als mehr oder weniger engen Bezug zu diesem Prototypen verstehen.

Eine zweite, für die Prototypentheorie wichtige Beobachtung ist, dass Kategorisierungen auch dann noch funktionieren, wenn die semantischen Merk-

male, die man zuweisen möchte, nicht oder jedenfalls nicht vollständig zutreffen. Man kann von einer Versuchsperson die Auskunft bekommen, dass der Unterschied zwischen „Kanne“ und „Krug“ darin besteht, dass die Wanddicke *und(!)* der Inhalt entscheidend sind, und wenn nicht die Wanddicke, dann der Inhalt, oder umgekehrt. Das bedeutet eine gewisse Variabilität der Merkmalsdefinition. (Einführend dazu Kleiber, 1998: 18 ff.)

Kategorien, die Eigenschaften zeigen, die den beiden Beobachtungen entsprechen, werden im Folgenden (in gewissem Sinne abgekürzt) als „prototypisch“ bezeichnet, diese Bezeichnung wird also nicht für den Prototypen einer Kategorie reserviert. Solche Kategorien haben dann die Eigenschaft der „Prototypizität“.

Man könnte nun versucht sein, neuronale Konstruktionen einzuführen, die speziell diese beiden Beobachtungen erklären könnten, z. B. einen Mechanismus, der eine Art Vergleich mit einem Prototypen leisten würde, also im Beispiel „Obst“ den Vergleich eines Inputs mit dem Prototypen „Apfel“, wohlgerichtet, mit dem Ziel einer Kategorisierung nicht als APFEL, sondern als OBST. Solche Vergleichsoperationen sind in einer symbolischen Architektur wahrscheinlich durchaus denkbar, es ist aber völlig unklar, wie dergleichen in einer neuronalen Architektur und in der erforderlichen Häufigkeit (das heißt für sehr viele Konzepte) zu realisieren wäre. Schon aus diesem Grund lohnt es sich, Lernvorgänge für Kategorien und die Anwendung erworbener Konzepte noch einmal etwas genauer anzusehen, in der Hoffnung, dass sich aus der Struktur der Lernergebnisse Hinweise für eine Lösung des Problems ergeben.

An dem Erwerb des Konzepts eines Alltagsgegenstandes, wie es z. B. eine Kanne ist, sind nicht nur einige wenige und gleichbleibende Wahrnehmungskomponenten, sondern eine größere und wechselnde Zahl unterschiedlicher primitiverer Konzepte beteiligt. Es muss auch angenommen werden, dass sich der Lernprozess über einen gewissen Zeitraum erstreckt, in dem sich Wahrnehmungen, die den Gegenstand betreffen, zwar teilweise wiederholen, aber nicht konstant vorhanden sind. Diese Voraussetzungen sind in der folgenden Simulation nachgebildet.

Die Architektur enthält eine Reihe von Großmutterzellen, die primitivere Konzepte realisieren. Für diese Konzepte wird auf die Darstellung des Erwerbsvorgangs, der Repräsentationsdetails und damit auch der Redundanz verzichtet und sie erhalten einen externen Input. Sie gelten trotzdem nicht als atomar, das heißt, es ist angenommen, dass sie über Bahnen erreicht werden, die sie letztlich mit Sinneswahrnehmungen oder auch Gedächtnisinhalten (z. B. für Szenen) verbinden. Die Bedeutung der in der Simulation verwendeten Zellidentifikationen kann man sich etwa wie in der folgenden

Tabelle vorstellen (alle Merkmale sind positiv spezifiziert zu denken, das Nicht-Vorhandensein eines Merkmals ist neuronal, jedenfalls in einer Struktur, die Lernvorgänge ermöglicht, nicht darstellbar):

kaf = mit Kaffee gefüllt
tee = mit Tee gefüllt
bau = bauchige Form
tul = röhrenförmiger Ausgießer (Tülle)
por = aus Porzellan
met = aus Metall (z. B. Messing)
wss = weiß
mus = gemustert
dec = mit Deckel
hss = heiß
fls = für Flüssigkeiten
hol = Hohlraum

Der dadurch gebildeten Eingabeschicht steht ein Block von 200 undefinierten Großmutterzellen gegenüber, jede Zelle der Eingabeschicht hat 15 Verbindungen mit beliebig ausgewählten Zellen des undefinierten Blocks. Diese Verbindungen sind zunächst schwach (2% des Werts, der bei Erregung zum Feuern der Zielzelle führen würde), sie werden über den Input durch Lernvorgänge verstärkt. Alle übrigen das Zellverhalten bestimmenden Parameter sind nach sehr groben Kriterien und ohne große Skrupel festgelegt: Die Abschwächung der Synapsen in der Zeit bei fehlender Erregung ist relativ niedrig eingestellt, die Lernrate pro Erregung einer Verbindung relativ hoch. Der Einfluss der Lernrate wird unten noch etwas genauer diskutiert.

Diese – zugegebenermaßen etwas unrealistischen – Einstellungen ermöglichen es, schon durch relativ wenige Eingabevorgänge brauchbare Ergebnisse zu erzielen. Die Eingabe besteht in dem folgenden Datensatz:

kaf 1 *bau* 2 *tul* 3 *por* 3 *wss* 4 *mus* 4 *hss* 5
kaf 101 *bau* 102 *tul* 103 *por* 103 *wss* 104 *mus* 104 *hss* 105
kaf 201 *bau* 202 *tul* 203 *por* 203 *wss* 204 *mus* 204 *hss* 205

hol 1001 *bau* 1002 *tul* 1003 *por* 1003 *wss* 1004 *mus* 1004 *dec* 1005
hol 1101 *bau* 1102 *tul* 1103 *por* 1103 *wss* 1104 *mus* 1104 *dec* 1105
hol 1201 *bau* 1202 *tul* 1203 *por* 1203 *wss* 1204 *mus* 1204 *dec* 1205

Das heißt, die Zelle *kaf* erhält im Zeittakt 1 einen Input, der bewirkt, dass sie im folgenden Zeittakt ein Aktionspotenzial produziert. Für *bau* geschieht

das in Zeittakt 2 usw., zwischen dem ersten und dem zweiten Eingabeblock verstreicht, wenn man den Zeittakt mit einer Millisekunde veranschlagt, eine Zeitspanne von einer Sekunde.

Simulation:
 Prototypizität I, Architekturdarstellung
 Ergänzender Hinweis: Steuerung über die Leertaste ergibt den besten Eindruck vom Simulationsverlauf.

Wenn man die Simulation verfolgt, fällt auf, dass sehr viele der 200 zunächst undefinierten Zellen am Ende Aktionspotenziale abgeben. Die betreffenden Zellen sind also als Großmutterzellen mit einer festen Bedeutung etabliert. Die Abbildung 2.4.5–1 zeigt das für den Zeittakt 1208.

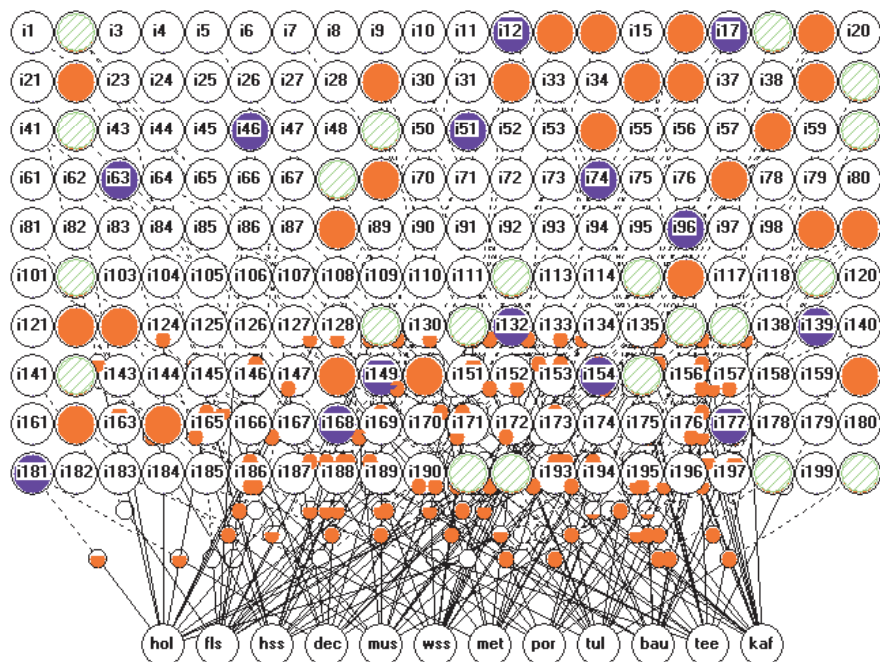


Abbildung 2.4.5–1: Simulation zum Lernen semantischer Kategorien. Simulationszustand nach Eingabe von zwei unterschiedlichen Merkmalsätzen. Ganz (rot) ausgefüllte Zellen produzieren ein Aktionspotenzial, für (grün) schraffierte Zellen liegt das maximal zwei Millisekunden zurück, sie sind in der Refraktärphase.

Die meisten der definierten Großmutterzellen sind einfach Repräsentanten einzelner Eingabemerkmale, also sozusagen Kopien der Eingabezellen. Einige Zellen repräsentieren aber auch Bündel von Eingabemerkmale. Diese Zellen zusammen mit den Merkmalskombinationen, auf die sie reagieren und den jeweiligen Verbindungsgewichten, auf einen Schwellenwert von 50 bezogen, sind in der folgenden Tabelle aufgeführt (ganz schwache zusätzliche Verbindungen, deren Effektivität sich gegenüber dem Ausgangswert nicht nennenswert verändert hat, sind nicht aufgenommen):

<i>i2</i>	<i>tul</i> 38.78 <i>por</i> 38.78
<i>i5</i>	<i>bau</i> 38.78 <i>bau</i> 38.78
<i>i12</i>	<i>kaf</i> 38.78 <i>wss</i> 38.78
<i>i17</i>	<i>kaf</i> 38.78 <i>wss</i> 38.78
<i>i29</i>	<i>bau</i> 19.45 <i>bau</i> 19.45 <i>mus</i> 19.45
<i>i58</i>	<i>bau</i> 38.78 <i>wss</i> 38.78
<i>i95</i>	<i>kaf</i> 38.78 <i>hss</i> 38.78
<i>i99</i>	<i>bau</i> 38.78 <i>mus</i> 38.78
<i>i100</i>	<i>por</i> 38.78 <i>wss</i> 38.78
<i>i119</i>	<i>bau</i> 38.78 <i>tul</i> 38.78
<i>i135</i>	<i>kaf</i> 38.78 <i>hss</i> 38.78
<i>i137</i>	<i>tul</i> 38.78 <i>por</i> 38.78
<i>i142</i>	<i>tul</i> 38.78 <i>por</i> 38.78
<i>i149</i>	<i>mus</i> 38.78 <i>hss</i> 38.78
<i>i150</i>	<i>por</i> 38.78 <i>wss</i> 38.78
<i>i155</i>	<i>bau</i> 19.45 <i>tul</i> 19.45 <i>tul</i> 19.45 <i>por</i> 19.45 <i>por</i> 19.45
<i>i177</i>	<i>por</i> 38.78 <i>hss</i> 38.78
<i>i181</i>	<i>dec</i> 38.74 <i>hol</i> 38.74

Bis auf eine Ausnahme bestehen alle gebildeten komplexeren Kategorien aus Kombinationen von zwei Eingabemerkmale. Das ist durch die relativ geringe Anzahl von Verbindungen der Eingabezellen mit dem Satz der undefinierten Zellen zu erklären. Es ist offen, wie hoch die Konnektivität im biologischen Original anzusetzen ist. Wie zu erwarten war, sind es beliebige Kombinationen, es gibt aufgrund des beschränkten Eingabematerials keine Häufigkeitseffekte in dem Sinn, dass ein über viele Wahrnehmungen relativ konstantes Merkmalsbündel hauptsächlich repräsentiert würde. Die Kategorie *i155* wird als einzige durch die Kombination von drei Merkmalen gebildet. (Die zufällige Bestimmung der Verbindungen hat zur Folge, dass eine einzelne Eingabezelle auch zwei Verbindungen mit einer undefinierten Zelle haben kann, diese doppelten Verbindungen sind gleichbedeutend mit stärkeren Einzelverbindungen, vgl. auch *i29*). Diese Kategorie zeigt eine weitere wichtige Eigenschaft: Es sind eigentlich nicht drei Eingabemerkmale

erforderlich, um die Großmutterzelle *i155* überschwellig zu erregen, sondern es genügen beliebige Kombinationen von zwei der konstitutiven Merkmale.

Man kann sich vorstellen, dass, eine Erhöhung der Zahl verstärkbarer Verbindungen auf potenziellen Großmutterzellen vorausgesetzt, die Variabilität auch noch wesentlich größer werden kann. Es entstehen dann Kategorien, die jede für sich genommen eine Auswahl möglicher Merkmale anbietet, von denen eine ausreichende Zahl, in ausreichender Stärke, im Input vorhanden sein muss, damit die Kategorie aktiviert werden kann. Die Kategorie bekommt damit eine gewisse „Unschärfe“. Wohlgemerkt: Das heißt nicht, dass die Kategorien fließend ineinander übergehen. Die Tasse wird nicht (notwendig) zu einer „schlechten“ Kanne. Wohl aber ist es möglich, ein Exemplar der Kategorie zuzuordnen, auch dann, wenn es nicht einer starren Definition genügt. Die Kategorie *i155* entspricht damit den Beobachtungen, die zu Anfang dieses Kapitels bezüglich der Kategorisierung festgestellt worden sind. Sie ist als prototypische Kategorie zu interpretieren. Die Prototypizität entsteht aufgrund prinzipieller Eigenschaften des neuronalen Lernprozesses und braucht keine zusätzlichen Mechanismen. Auch der Kategorisierungsprozess aufgrund gelernter Kategorien ist maximal einfach und setzt keine Vergleichsoperation mit einem herausgehobenen Exemplar voraus.

An dieser Stelle ist es interessant, noch etwas genauer zu überprüfen, unter welchen Bedingungen eine Prototypizität der beschriebenen Art tatsächlich zustande kommt.

In den folgenden Simulationen sind, um die Handhabung zu erleichtern, die auf die Zelle *i155* einwirkenden Input-Zellen und die Zelle *i155* selbst isoliert. Es wird zunächst gezeigt, dass die Festsetzung der Lernrate einen Einfluss auf die Prototypizität der Kategorie hat. Anschließend wird dargestellt, was sich ergibt, wenn die Zahl der Eingänge auf die Zielzelle erhöht wird und wenn die Lernraten eine gewisse Variabilität aufweisen.

Simulationen:

Lernen einer prototypischen Kategorie:

Erster Versuch: Lernrate 21.5.

Zweiter Versuch: Lernrate 10.75.

Dritter Versuch: Lernrate 6.5.

Vierter Versuch: Lernrate 6.5, vier Eingänge.

Fünfter Versuch: Streuung der Lernraten, vier Eingänge.

(Lernraten bezogen auf einen Schwellenwert von 50.)

Ergänzender Hinweis: Die Verbindungsgewichte am Ende der Simulation müssen an einem Zustandsprotokoll, das über „Datei – Zustand speichern“ erzeugt worden ist, abgelesen werden.

Wie die folgende Tabelle zeigt, bleiben zwar die Verbindungsgewichte in einem weiten Bereich von Lernraten nahezu gleich, das ändert sich aber, wenn die Lernrate so gering wird, dass mehr als drei Inputs gebraucht werden, um die Zielzelle überschwellig zu erregen. Man kann ohne weiteres abschätzen, dass das Phänomen damit aber letztlich nur verschoben ist. Wenn mehr als vier Eingabeeinheiten auf die Zielzelle einwirken, reicht auch die niedrige Lernrate aus, um wieder eine prototypische Kategorie zu bilden. Schließlich wird das Bild auch dadurch bunter, dass man mit einer gewissen Streuung der Lernraten (und natürlich auch anderer Zellparameter) rechnen muss.

Lernrate	Verbindungsmuster
21.5	<i>bau</i> 19.45 <i>tul</i> 19.45 <i>tul</i> 19.45 <i>por</i> 19.45 <i>por</i> 19.45
10.75	<i>bau</i> 19.44 <i>tul</i> 19.44 <i>tul</i> 19.44 <i>por</i> 19.44 <i>por</i> 19.44
6.5	<i>bau</i> 11.79 <i>tul</i> 11.79 <i>tul</i> 11.79 <i>por</i> 11.79 <i>por</i> 11.79
Einführung einer zusätzlichen Verbindung:	
6.5	<i>kaf</i> 5.88 <i>bau</i> 11.79 <i>tul</i> 11.79 <i>tul</i> 11.79 <i>por</i> 11.79 <i>por</i> 11.79
Lernraten gestreut:	
5.4 bis 21.5	<i>kaf</i> 4.90 <i>bau</i> 19.44 <i>tul</i> 9.82 <i>tul</i> 9.82 <i>por</i> 9.82 <i>por</i> 38.78

Die Halbierung der Lernrate hat, wie gezeigt worden ist, keinen nennenswerten Effekt auf die Formierung der Kategorie *i155*. Wohl aber ändert sich die Menge der gebildeten komplexen Kategorien insgesamt relativ deutlich, wie in der folgenden Simulation gezeigt wird, die wieder das gesamte Netz verwendet, nicht nur die herauspräparierte Kategorie.

Simulation:

Prototypizität II, Halbierung der Lernrate

Ergänzender Hinweis: Steuerung über die Leertaste ergibt den besten Eindruck vom Simulationsverlauf.

Es entstehen mehr komplexe Kategorien. Gleichzeitig nimmt die Zahl der Kategorien, die durch einzelne Eingabemerkmale überschwellig erregt werden können, ab. Einzelne Merkmale bilden bei dem verwendeten Input nur noch dann funktionsfähige Kategorien in dem Block der undefinierten Zellen, wenn sie (mindestens) zwei Verbindungen auf eine Zelle haben. Da der unterschwellige Wert des EPSP einiger Zellen aber über 48 (bei einer Schwelle von 50) ist, würde eine Fortsetzung des Inputs mit denselben Werten sehr rasch wieder auf einige Kopien einzelner Merkmale führen. Zellen mit mehr Verbindungen werden allerdings in jedem Fall begünstigt. Die Abbildung 2.4.5-2 zeigt den Simulationszustand zu dem selben Zeittakt wie Abbildung 2.4.5-1.

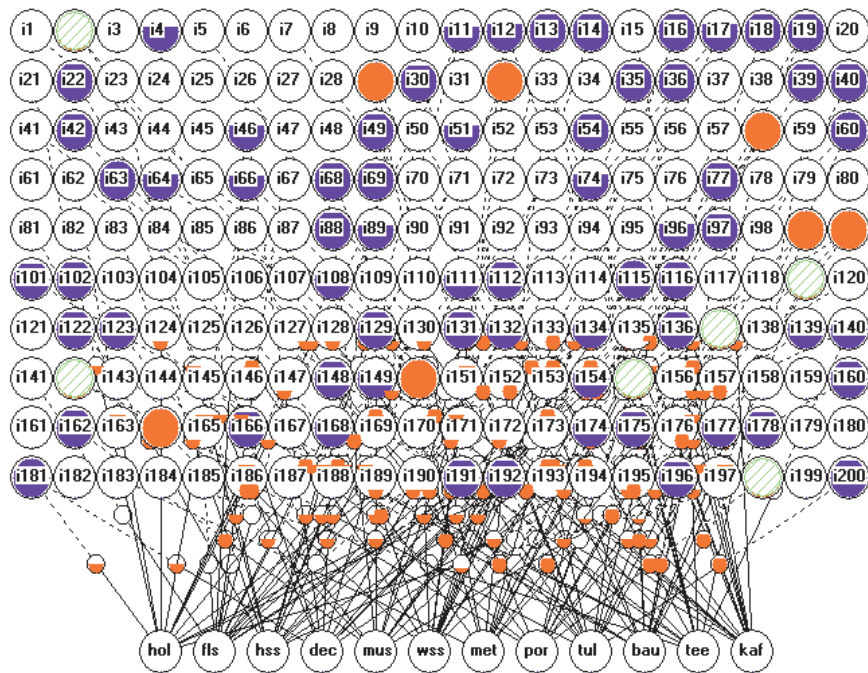


Abbildung 2.4.5-2: Simulation zum Lernen semantischer Kategorien. Zustand im gleichen Zeittakt wie in Abbildung 2.4.5-1, reduzierte Lernrate.

Die Tabelle der komplexen Kategorien (Verbindungsgewichte wieder auf einen Schwellenwert von 50 bezogen) zeigt, dass die bei der höheren Lernrate gebildeten Kategorien beibehalten sind. Die größere Zahl entsteht durch Erhöhung der Redundanz (im Fall *dec* + *hol*) und durch die Repräsentation zusätzlicher Merkmalskombinationen.

i2 tul 29.10 por 29.10
i11 hss 28.80 hol 29.03
i12 kaf 29.09 wss 29.10
i17 kaf 29.09 wss 29.09
i29 bau 19.44 bau 19.44 mus 19.44
i58 bau 29.10 wss 29.10
i60 tul 48.18 dec 9.75
i64 kaf 28.80 dec 29.04
i74 dec 29.04 hol 29.03
i95 kaf 29.09 hss 29.10

<i>i99</i>	<i>bau</i> 29.10	<i>mus</i> 29.10
<i>i100</i>	<i>por</i> 29.10	<i>wss</i> 29.10
<i>i111</i>	<i>hss</i> 28.80	<i>hol</i> 29.03
<i>i115</i>	<i>por</i> 48.18	<i>dec</i> 9.75
<i>i119</i>	<i>bau</i> 29.10	<i>tul</i> 29.10
<i>i134</i>	<i>hss</i> 28.80	<i>hol</i> 29.03
<i>i135</i>	<i>kaf</i> 29.09	<i>hss</i> 29.10
<i>i137</i>	<i>tul</i> 29.10	<i>por</i> 29.10
<i>i142</i>	<i>tul</i> 29.10	<i>por</i> 29.10
<i>i149</i>	<i>mus</i> 29.10	<i>hss</i> 29.10
<i>i150</i>	<i>por</i> 29.10	<i>wss</i> 29.10
<i>i155</i>	<i>bau</i> 19.44	<i>tul</i> 19.44
	<i>tul</i> 19.44	<i>por</i> 19.44
	<i>por</i> 19.44	<i>por</i> 19.44
<i>i177</i>	<i>por</i> 29.10	<i>hss</i> 29.10
<i>i181</i>	<i>dec</i> 29.07	<i>hol</i> 29.06
<i>i198</i>	<i>por</i> 48.18	<i>hol</i> 19.42

Im Simulationsmodell kann man einen Maximalwert für die Leistungsfähigkeit von Synapsen angeben und damit gewährleisten, dass nicht nur vorübergehend, sondern grundsätzlich Großmutterzellen nur durch die Kombination einer gewissen Anzahl von Eingängen gültig definiert werden. Im biologischen Original sind die erreichbaren Synapseneffektivitäten wahrscheinlich ebenfalls begrenzt und man kann sich fragen, in welchem Umfang eine monosynaptische Erregung überhaupt möglich ist.

Wenn man sich vorstellt, dass gleichzeitig mit der Wahrnehmung der Merkmalsbündel mit einer ausreichenden Wiederholungsrate der Ausdruck „Kanne“ wahrgenommen wird, also ein Lernvorgang (der an dieser Stelle noch nicht besprochen werden kann) zu einer Zuordnung des Ausdrucks mit den aktivierten Konzepten führt, ergibt sich eine für die Diskussion innerhalb der Prototypentheorie interessante Folgerung: Wir bekommen die Möglichkeit, zwischen Polysemie und Prototypizität klar zu unterscheiden. Polysemie bedeutet (definitionsgemäß) die Zuordnung verschiedener Kategorien zu einer lexikalischen Ausdrucksseite. Prototypizität betrifft dagegen eine einzelne Kategorie, die eine sozusagen „interne“ Variabilität aufweist. Die in den Simulationsexperimenten gebildeten Kategorien können allesamt mit derselben Ausdrucksseite „Kanne“ verbunden sein. Die Benennung „Kanne“ kann dann ausgelöst werden durch irgendeine der etablierten Merkmalskombinationen. Auch das, nicht nur die Prototypizität der einzelnen prototypischen Kategorie, ergibt besondere sprachliche Leistungen, und zwar der Art, dass sie nur schwer von denen der Prototypizität im engeren Sinne unterscheidbar sind. Der Übergang zu einer auch in der Lexikographie notierten Polysemie ist fließend.

Das wichtigste Fazit aus den vorangegangenen Überlegungen ist, dass die Entstehung prototypischer Kategorien als ein Abfallprodukt unverzichtbarer neuronaler Lernmechanismen angesehen werden muss. Prototypische Kategorien verleihen dem Wahrnehmungsprozess eine besonders wertvolle Flexibilität. Es ist ein bemerkenswertes Produkt der Evolution, dass die Lernmechanismen, die nicht erst beim Menschen gelten, einen solchen flexiblen Wahrnehmungsprozess ermöglichen. Trotz ihrer Variabilität haben prototypische Kategorien scharfe Grenzen: Eine Salatschüssel ist keine Kanne, selbst wenn man durch geschicktes Fragen herausbekommen kann, dass eine Salatschüssel eine bessere Kanne ist als ein Stuhl. Die klassische Prototypentheorie ist eine Art Gratwanderung zwischen außersprachlicher Kategorisierung und linguistischer Semantik. Wenn man aber auf eine eigenständige linguistische Semantik verzichtet, werden die Dinge klar: Prototypizität entsteht zunächst sprachfern. Die einzelsprachliche Anpassung der Kategorien ist eine besondere Leistung des Spracherwerbsprozesses (siehe Abschnitt 2.4.4).

Der für die Prototypentheorie zentrale Bewertungsvorgang in Laborsituationen bleibt allerdings vorerst ungeklärt und kann erst in Abschnitt 2.4.8 genauer beleuchtet werden. Nicht mit erfasst ist auch der Fall, dass durch die Bildung von Vorstellungen „defekte“ oder abweichende Exemplare einer Kategorie doch noch eingeordnet werden können. Die Bildung von Vorstellungen ist Gegenstand des Kapitels 2.5.

2.4.6 Kontexte

Der Begriff „Kontext“ wird in so verschiedener Weise gebraucht, dass es nicht lohnt, sich mit den Positionen in der diesbezüglichen Literatur detailliert auseinanderzusetzen. Stattdessen soll hier eine Vereinbarung gelten: „Kontext“ meint im Folgenden eine Informationsmenge, die einen zu einem bestimmten Zeitpunkt stattfindenden mentalen Verarbeitungsprozess beeinflusst. Er stellt einen Hintergrund dar, auf dem ein jeweils vordergründiger Prozess abläuft. Dieser Hintergrund ist ständiger Veränderung unterworfen, indem die gegenwärtige Verarbeitungssituation jeweils den Hintergrund für eine folgende Verarbeitungssituation abgibt („sequenzieller Kontext“). Vor allem diese Funktion von Kontext soll hier behandelt werden. Man kann aber natürlich auch eine Ansammlung aktuell verfügbarer Daten als Kontext bezeichnen. Semantische Merkmale, die zu einem bestimmten Zeitpunkt zur Verfügung stehen, sind ein Beispiel dafür („punktueller Kontext“). Das Problem bei der Bestimmung von „Kontext“ ist in erster Linie, zu einer genügend engen Abgrenzung dessen zu kommen, was als Kontext gelten

soll. In dieser Hinsicht ist die hier gewählte Ausdehnung minimal: Kontext ist nur, was *im Moment* kausal bestimmend wirkt, nicht z. B. etwas, was zu einem laufenden Verarbeitungsprozess vielleicht auch noch hinzugezogen werden kann.

Die Verarbeitung sequenzieller Kontexte ist ein prinzipielles Charakteristikum natürlicher intelligenter Prozesse. Eine Idee zum Funktionieren von Kontexten ist, dass vorangegangene Ereignisse jeweils Erwartungen bezüglich der folgenden Ereignisse erzeugen. Ein klassisches anschauliches Beispiel für diese Sicht – mit Bezug auf die Frame-Theorie in der künstlichen Intelligenz – findet sich bei Benjamin J. Kuipers:

„Consider for a moment an intuitive description of how a frame system might work in the everyday vision process. As you are walking through an unfamiliar house, you come to a normal interior-type door, open it, and walk through. At the moment that you open the door, your (entirely reasonable) expectations have already brought a “room” frame to mind. There is no delay in comprehending the fact that you see four walls, floor, and ceiling, since you already “knew” that they would be there, even without having seen them. Indeed, if these expectations had not been fulfilled, and you had been presented with, say, a seashore instead, you would experience a sense of disorientation.“ (Kuipers, 1975: 154)

Die Erwartungen sind da, ehe ein folgendes Ereignis, dessen Interpretation sie beeinflussen, eintritt. Eine wichtige Beobachtung ist, dass diese Erwartungen enttäuscht werden können und dann ein Gefühl der Desorientierung auslösen. Man kann an dem Beispiel von Kuipers besonders schön sehen, dass Kontextverarbeitung eine wichtige Funktion hat bei der Wahrnehmung des Zusammenhangs in der Welt ganz allgemein.

Eine alternative Idee wäre, dass nicht Erwartungen erzeugt werden, die sozusagen den Ablauf eines künftigen Prozesses schon vorwegnehmen, sondern dass jeder Prozess, der zu einem bestimmten Zeitpunkt abläuft, auf Informationen zurückgreift, die durch zurückliegende Prozesse produziert worden sind. In diesem Fall gibt es keine sozusagen in die Zukunft gerichteten Erwartungen, sondern stattdessen in die Vergangenheit gerichtete Zugriffe auf vorausliegende Daten. Wenn Kontexte als Erwartungen zu verstehen sind, sind sie auf der Seite der aktuellen Vorgänge angesiedelt, wenn sie durch Rückgriff auf zurückliegende Informationen entstehend denkt, sind sie bei den Repräsentationen vergangener Prozesse angesiedelt.

Wenn von Zugriffen auf vorausliegende Daten die Rede ist, liegt es nahe, an symbolverarbeitende Architekturen zu denken, in denen von Daten und

deren Adressierung gesprochen werden kann (vgl. Kapitel 2.2.2). In einem neuronalen Modell könnte man folgende Konstruktion erwägen: Prozesse hinterlassen Spuren in Form von unterschwelligen EPSPs in den Netzbereichen, in denen sie stattgefunden haben. Nachfolgende Prozesse haben Verbindungen auf die Zellen, die die Spuren enthalten und haben damit die Möglichkeit, eine Rückmeldung aus diesen Zellen zu erzeugen, deren Auswertung dann zur Steuerung (z. B. zur Beseitigung der Mehrdeutigkeit eines Inputs) verwendet wird (Abbildung 2.4.6-1).

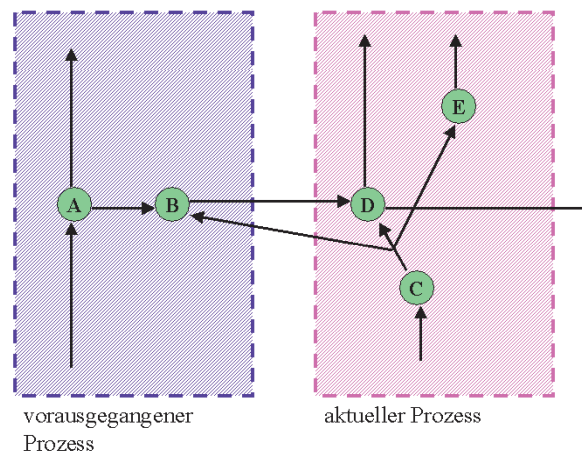


Abbildung 2.4.6-1: Architekturskizze zur Auswertung eines zurückliegenden Kontexts.

Es braucht keine komplizierten Überlegungen, um einzusehen, dass die Zelle *B* in Abbildung 2.4.6-1 eigentlich überflüssig ist. Von *A* ausgehend könnte direkt *D* mit einer Spur versehen werden, die dann in dem aktuellen Prozess zur Auswahl der entsprechenden, von *C* ausgehenden Bahn führen würde. Was damit aber entsteht, ist eine Architektur, die mit Erwartungen arbeitet, also mit Spuren vorangehender Prozesse innerhalb eines aktuellen Bereichs. Es wäre kaum zu verstehen, wenn die Evolution diese Vereinfachungsmöglichkeit bildlich gesprochen „übersehen“ hätte.

Für die Repräsentation von Erwartungen (wenn wir uns für diese Variante entscheiden) gibt es verschiedene wichtige Bedingungen:

- Die Repräsentationen sind über Verbindungen aufzubauen, die durch Lernvorgänge etabliert werden müssen. (Der „Inhalt“ einer Erwartung besteht in der Auswahl der Verbindung, die benutzt wird.)

- Die Spuren (nicht die Verbindungen) müssen in nach Zehntelsekunden oder noch kleineren Einheiten zu bemessenden Zeiträumen zuverlässig aufgebaut und abgebaut werden.
- Wenn Zellen, die solche Erwartungen repräsentieren, eine steuernde Funktion haben sollen, müssen sie in den Verlauf des Prozesses, den sie steuern sollen, eingeschaltet sein.
- Auf ein längeres Stehenbleiben der Spuren kann verzichtet werden, wenn die Idee akzeptiert wird, dass Kontexte zeitlich unmittelbar wirksam sind. Andererseits muss doch eine gewisse zeitliche Flexibilität gewährleistet sein, der Zeitpunkt des Eintreffens des erwarteten Ereignisses kann nicht auf die Millisekunde genau festgelegt werden.

Unter diesen Bedingungen scheidet die Möglichkeit einer synaptischen Repräsentation der aktuellen Spuren selbst (Repräsentation durch Veränderung der Effektivität von Synapsen) sicher aus. Es bleibt nur die Möglichkeit, unterschwellige EPSPs zu verwenden. Ein Prozess, der Erwartungen erzeugt, hinterlässt unterschwellige EPSPs in Zellen, die zur Repräsentation möglicher Folgeprozesse gehören. Bahnen innerhalb der Repräsentation der Folgeprozesse, die über Zellen führen, die keine solchen „Erwartungspotenziale“ aufweisen, werden blockiert. Bahnen, die über Zellen mit Erwartungspotenzialen führen, sind durch diese Erwartungspotenziale sozusagen freigeschaltet. Ungewöhnliche Folgeerscheinungen wie der Strand hinter der Zimmertür in dem oben zitierten Beispiel von Kuipers gehören nicht zu den freigeschalteten Netzbereichen, die Verarbeitung des ungewöhnlichen Inputs stockt und führt zu einer Inkohärenzreaktion.

EPSPs, die nicht schließlich die Bildung eines Aktionspotenzials auslösen, klingen innerhalb eines mehr oder weniger großen Zeitfensters ab. Die Verwendung einer mit einem unterschweligen EPSP versehenen Zelle als Schalter setzt voraus, dass dieses EPSP durch einen weiteren Input, der ein „Abfragepotenzial“ erzeugt, über den Schwellenwert hinaus ergänzt wird. Wenn der Verlauf der Abnahme eines unterschweligen Potenzials in der Zelle als annähernd exponentiell beschrieben werden kann, aber auch unabhängig von dieser Annahme, gilt, dass eine sichere Funktion einschließlich einer gewissen zeitlichen Flexibilität nur dann erzielt wird, wenn das Abfragepotenzial generell vergleichsweise hoch ist. Eine flexible Anpassung dieses Potenzials in Abhängigkeit von dem Erregungszustand der Zelle, die man sich als Alternative prinzipiell denken könnte, wird man nicht annehmen wollen.

Man beachte jetzt, dass die Idee eines Abfragepotenzials notwendig impliziert, dass dieses Potenzial auch bei Zellen erzeugt wird, die *nicht* durch ein

Erwartungspotenzial vorbereitet sind. Im Beispiel des unerwarteten Strands hinter der Zimmertür sind es die Zellen, den den Strand repräsentieren, in denen ein Abfragepotenzial entsteht. Im Beispiel der Verarbeitung eines mehrdeutigen Inputs werden *alle* Bedeutungen abgefragt, nicht nur die durch den Kontext begünstigten. Die Konsequenz ist, dass überall dort, wo die Abfrage nicht zum Entstehen eines Aktionspotenzials geführt hat, relativ hohe erregende Potenziale stehen bleiben, für die dann dieselben Abnahmebedingungen gelten, wie für die Erwartungspotenziale, und die damit von Erwartungspotenzialen nicht unterschieden werden können. Eine sinnvolle und verlässliche Kontextsteuerung kann auf diese Weise nicht entstehen.

Die Konsequenz ist, dass man nach Möglichkeiten suchen muss, die Abfragepotenziale, die sozusagen keinen Erfolg gehabt haben, möglichst rasch (und das muss hier heißen: aktiv) wieder abzubauen. Dergleichen kann im Nervensystem nur durch die Wirkung hemmender Zellen erzielt werden. Es ergibt sich die – vorläufige(!) – Konstruktion der Abbildung 2.4.6–2:

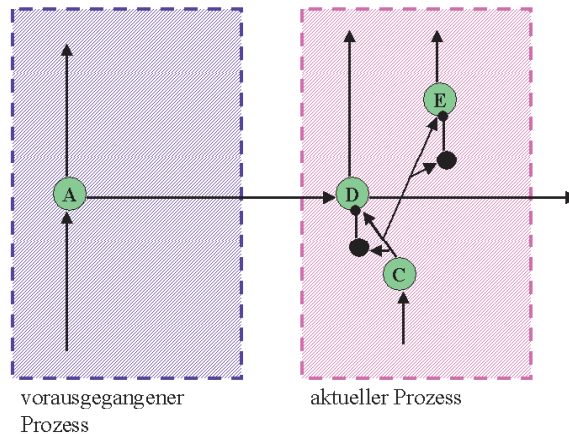


Abbildung 2.4.6–2: Architekturskizze zur Auswertung von Erwartungspotenzialen. Die größeren schwarz ausgefüllten Kreise stellen hemmende Zellen dar. Bei diesen Zellen werden die Pfeilspitzen der Axone durch kleine Kreisscheiben ersetzt.

Diese Konstruktion ist, von den geforderten Leistungen her gesehen, für die Kontextverarbeitung ausreichend. Man muss aber zusätzlich beachten, dass einige der dabei verwendeten Verbindungen über Lernprozesse etabliert gedacht werden müssen. Es muss gezeigt werden, dass solche Prozesse möglich sind oder noch besser, wie solche Prozesse aussehen.

Betroffen sind in Abbildung 2.4.6–2 alle Verbindungen, die zu den Zellen *D* und *E* führen. Das größte Problem stellen dabei wohl die hemmenden Verbindungen dar. Es muss ja nicht nur einfach eine benutzte (ein Aktionspotenzial führende) Verbindung verstärkt werden, sondern die Verstärkung muss auf Verbindungen mit ganz bestimmten Zellen beschränkt bleiben, wobei die Auswahl nicht einmal so geschehen kann, dass alle gerade unterschwellig erregten Zellen gültige Zielzellen sind. Ein zweites Problem stellen die „waagrechten“ Verbindungen dar, die die Erwartungspotenziale erzeugen.

Erwartungspotenziale sind unterschwellige Potenziale. Die längerdauernde Verstärkung der entsprechenden Synapsen muss an das Feuern der Zielzelle gebunden sein. Diese Zielzelle sollte eine erwartete Kategorie repräsentieren. Wenn man annehmen möchte, dass diese Kategorie schon existiert, ehe sie in einen Erwartungszusammenhang eingebunden wird, ist eine Verstärkung der Verbindung für das Erwartungspotenzial nicht mehr möglich, da die entsprechende Großmutterzelle nach dem Feuern (ausgelöst durch die Inputs, die die Kategorie bilden) die weitere Verstärkung von Synapsen nicht mehr unterstützen darf. Wenn man, um diese Schwierigkeit zu vermeiden, annimmt, dass Kontextverbindung und Kategorienbildung gleichzeitig erfolgen, bekommt man Kategorien, die nicht in mehreren Kontexten wiederverwendet werden können, was mit der alltäglichen Erfahrung wohl kaum zu vereinbaren ist.

Alle beschriebenen Lernprobleme (und einige andere, die allerdings an dieser Stelle nur schwer zu erläutern wären) verschwinden, wenn man eine Unterscheidung trifft zwischen Verbindungen, die funktionsfähig und fest sind, also nicht durch Lernprozesse verstärkt werden müssen, und solchen, die vor einem Lernprozess schwach sind und durch Lernen verstärkt werden können. Die hemmenden Verbindungen sind, angesichts der oben beschriebenen Lernprobleme, besonders gute Kandidaten für feste Verbindungen. Wenn die hemmenden Verbindungen fest sind, liegt es nahe, auch für die parallel verlaufenden erregenden Verbindungen diese Annahme zu versuchen. Die gewünschte lernabhängige Kontextsteuerung, speziell das Angebot möglicher Alternativen, das heißt die Bindung verschiedener Konzepte an verschiedene Kontexte, kann dann aber nur noch so realisiert werden, dass zusätzliche lernfähige Zellen eingeführt werden, in der Abbildung 2.4.6–3 mit *d* und *e* bezeichnet. Die neu eingeführten Zellen repräsentieren dann die Konzepte bzw. Kategorien, die Zellen *D* und *E* sind nur noch für die Kontextanbindung verantwortlich.

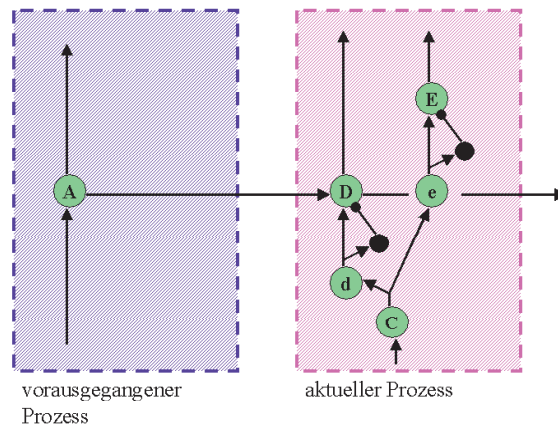


Abbildung 2.4.6-3: Architektur zur Auswertung von Erwartungspotenzialen. Entwurf nach der Beachtung von Lernproblemen.

Vielleicht etwas überraschend ist damit auch das Problem des Lernens der Kontextverbindung und das Problem der Wiederverwendbarkeit gelernter Kategorien in verschiedenen Kontexten gelöst, da die Zelle, die die Kategorie bildet, nicht mehr identisch ist mit der Zelle, die die Kontextverbindung bildet und die Zelle, die die Kontextverbindung bildet, aus diesem Grund nicht feuert, ehe die Kontextverbindung eine ausreichende Stärke hat.

Zur Veranschaulichung der Funktion kann die folgende Simulation verwendet werden (die Architektur entspricht der Abbildung 2.4.6-3).

Simulation:

Kontextverarbeitung

In Zeittakt 5 ff. gibt die Zelle *A* ein Aktionspotenzial ab, das zu einem Erwartungspotenzial in *D* führt.

In Zeittakt 55 ff. erfolgt ein Input über die Zelle *C*, nur die über *D* führende Bahn bleibt weiter aktiv, in *E* bleibt die Erregung unterschwellig und wird über den hemmenden Input wieder abgebaut. *D* ist zu diesem Zeitpunkt in der Refraktärphase, der hemmende Input bleibt dort wirkungslos.

An dieser Stelle ist es wichtig, eine Konsequenz zu beachten, die sich aus der Vorstellung einer andauernden, allgegenwärtigen Kontextsteuerungsfunktion ergibt (wobei der Einfluss so gedacht wird, dass nichtzutreffende, also nicht durch den Kontext gebahnte Alternativen blockiert werden). In der Simulation ist angenommen worden, dass ein Input in der Zelle *A* ein einzelnes Aktionspotenzial auslöst. Es ist vernachlässigt worden, dass natürlich

auch für diese Zelle ein entsprechender Kontext gegeben sein muss. Wenn ein entsprechender Kontext vorhanden ist, kann der Input ein einzelnes Aktionspotenzial auslösen. Das ist ausreichend in dem Fall, dass z. B. eine Wahrnehmung nur sehr kurze Zeit dauert, vielleicht gerade eine Zehntelsekunde. Ein längeres Andauern hat aber zur Folge, dass mehrere Inputimpulse erscheinen und damit auch der Kontext jeweils erneuert werden müßte. Es ist klar, dass die Erneuerung des Kontexteffekts für die Zelle *A* von einer Aktivität dieser Zelle ausgehen muss und nicht von einem vorangegangenen Ereignis. Das heißt, das der Zelle *A* entsprechende Ereignis muss sich selbst erwarten, oder technisch: Die Zelle *A* muss ein Erwartungspotenzial auf sich selbst erzeugen. Das setzt voraus, dass diese Zelle eine Verbindung auf sich selbst zurück hat. Analoges gilt für die Zellen *D* und *E*. Diese Konstruktion bringt an sich keine grundsätzliche Schwierigkeit mit sich, es ist aber anzunehmen, dass eine solche Verbindung von Zellen mit sich selbst nicht über Lernvorgänge aufgebaut werden kann und damit angeboren sein muss (solche Verbindungen könnten erst geschaltet werden, *nachdem* die betroffenen Zellen gefeuert haben, was den für diese Zellen gültigen Lernprinzipien widerspricht; eine genauere Diskussion findet sich in Kochendörfer, 2002:100).

2.4.7 Szenen

Die in der Überschrift dieses Abschnitts verwendete Bezeichnung „Szenen“ soll sich hier auf Strukturen von Folgen von Erwartungen beziehen. Auf die Erwartung ZIMMERTÜR kann die Erwartung FLUR folgen, auf die Erwartung FLUR die Erwartungen ZIMMERTÜR oder HAUSTÜR (mehrere gleichrangig vorgesehene Möglichkeiten), vielleicht auch LÄNGLICH usw., das heißt, Erwartungen können sich zu Strukturen zusammenschließen. Interessant wird es, wenn solche Erwartungsstrukturen als Ganzes Einheiten bilden, feststellbar z. B. daran, dass es eine sprachliche Bezeichnung für den Gesamtkomplex gibt. Vermutlich bilden Komplexe, die man in der Linguistik mit „Skripts“ beschrieben hat, solche Szenen. Das Standardbeispiel dafür ist das sog. „Restaurantskript“ von Schank & Abelson (1977). Man muss allerdings feststellen, dass Skripts einige Eigenschaften haben, die zu den folgenden Überlegungen nicht gut passen. Wir verzichten auf eine genauere Auseinandersetzung.

Ein Restaurantbesuch besteht aus einer Abfolge von Ereignissen, z. B. „Restaurant betreten“, „Bestellen“, „Essen“, „Bezahlen“ und „Restaurant verlassen“. Diese Abfolge von Ereignissen kann nun nicht in derselben Weise wie eine Menge von semantischen Merkmalen zur Bildung einer Kategorie RESTAURANTBESUCH dienen, wenigstens dann nicht, wenn man sie als

zeitliche Abfolge von Erwartungen betrachtet. Merkmale, die zur Kategorisierung dienen, müssen in einem kleinen Zeitfenster, also in diesem Sinne „gleichzeitig“ aktiviert sein. Es ist auch gleichgültig, in welcher Reihenfolge sie innerhalb dieses Zeitfensters erscheinen. Ein zweiter wichtiger Punkt ist, dass die Kategorie RESTAURANTBESUCH aktiviert sein muss während des gesamten Ablaufs der Einzelereignisse. Die Kategorisierung erfolgt offenbar nicht erst nach Abschluss der Ereignisserie. Wenn man eine Person sieht, die alleine an einem quadratischen Tisch sitzt und mit einer zweiten Person spricht, die neben dem Tisch steht, offenbar aufmerksam zuhört und sich auf einem kleinen Schreibblock Notizen macht, wird man die Kategorie RESTAURANTBESUCH und (vielleicht einige andere passende) aktivieren, obwohl nur ein Ausschnitt („Bestellen“) aus der charakteristischen Ereignisserie wahrgenommen wird. An diesem Beispiel ist auch sofort zu erkennen, dass die Kategorie BESTELLEN ihrerseits auf analoge Probleme führt.

Die Zelle, die die Einheit RESTAURANTBESUCH repräsentiert, muss damit folgende Eigenschaften haben:

- Sie muss eine Konvergenz zeitlich aufeinanderfolgender Ereignisse leisten, das heißt, sie muss Aktionspotenziale abgeben, solange eines der Ereignisse andauert.
- Sie muss spezifisch sein für die ihr zugeordnete Ereignisstruktur. Der Regen vor dem Fenster gehört nicht zur Einheit RESTAURANTBESUCH.
- Die Einheit darf nicht identisch sein mit einem der Teilereignisse.

Eine wichtige weiterführende Überlegung ist jetzt, dass die Einheit RESTAURANTBESUCH nicht angeboren sein kann, sondern durch Lernprozesse gebildet werden muss. Da einerseits alle Verbindungen auf ein Neuron, das die Einheit repräsentieren soll, nach dem Lernvorgang überschwellige Potenziale auslösen müssen, ohne dass sich die aufeinanderfolgenden Potenziale überlagern, kann das Feuern der Zelle nicht den Abschluß des Lernprozesses zur Folge haben; der für die Kategorienbildung zuständige Lernprozess, wie er bisher beschrieben worden ist, kann das gewünschte Ergebnis nicht gewährleisten. Andererseits muss die Zielzelle einer zu verstärkenden Verbindung immer in irgendeiner Weise ausgezeichnet sein, es dürfen nicht Verbindungen mit beliebigen Zellen dauerhaft(!) verstärkt werden. Zusammen mit der gängigen Vorstellung, dass die Langzeitpotenzierung etwas mit dem Feuern der Zellen zu tun hat, ist es letztlich nur die folgende Konstruktion, die allen Bedingungen genügt:

- Die zusammenfassende Zelle ist fest, das heißt funktionsfähig, mit einer Kombination von Zellen verbunden (genauer: mit der Zelle, die für die Einbindung des Kontexts zuständig ist), die eine Kategorie aufgrund eines Merkmalsbündels bildet (z. B. die Kategorie RESTAURANTBESUCH).
- Da nur bestimmte Verbindungen, nämlich solche mit anderen Konzepten in einem bestimmten Zeitfenster, verstärkt werden sollen, muss über das Feuern der zusammenfassenden Zelle ein solches Zeitfenster geöffnet werden.
- Die Verstärkung potentieller Verbindungen wird dann durch die Aktivität der Repräsentationen der einzelnen Teilereignisse bewirkt. Das Zeitfenster kann nicht größer sein, als für die Anbindung eines einzelnen Ereignisses erforderlich ist (sicher kleiner als eine Zehntelsekunde).
- Es ist zwar anzunehmen, dass auch die Einzelereignisse ihrerseits feste Verbindungen mit zusammenfassenden Zellen haben, die zu verstärkenden Verbindungen mit der übergeordneten Szene können aber nicht von diesen Zellen ausgehen, da sonst Ringbildungen mit hochfrequenter kreisender Erregung entstehen würden (das heißt, es würden ständig epileptische Anfälle ausgelöst).

Zusammen mit den in dem vorangegangenen Abschnitt angestellten Überlegungen zur Kontextverarbeitung führt das zu der in Abbildung 2.4.7–1 skizzierten Architektur.

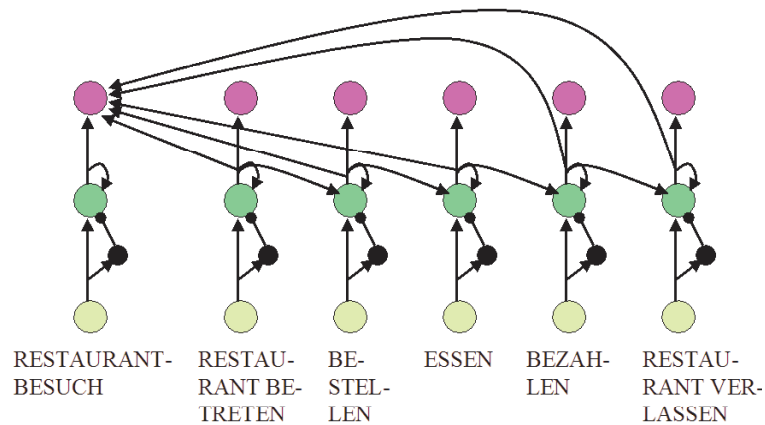


Abbildung 2.4.7–1: Architekturskizze zur Repräsentation einer Szene, am Beispiel des Restaurantbesuchs.

Die in Abbildung 2.4.7–1 verschieden eingefärbten Zellen haben unterschiedliche Eigenschaften, gehören also verschiedenen Zelltypen an. In Übereinstimmung mit früheren Publikationen (z. B. Kochendörfer, 2000 und 2002) werden für diese verschiedenen Zelltypen die in Abbildung 2.4.7–2 zugeordneten Bezeichnungen verwendet.

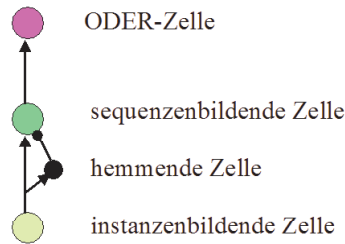


Abbildung 2.4.7–2: Zelltypen.

Bemerkungen zu den einzelnen Zelltypen:

Instanzenbildende Zellen. Die Bezeichnung erinnert daran, dass diese Zellen nicht nur Kategorien in der oben beschriebenen Weise bilden, sondern dass es auch möglich ist, dass im Verarbeitungsprozess entstandene Kategorien durch diese Zellen weiterentwickelt und auch vervielfältigt werden können. Über den Zusammenhang mit der Funktion der ODER-Zellen wird gleich noch zu reden sein.

Hemmende Zellen. Es ist biologisch nicht möglich, dass Zellen gleichzeitig erregende und hemmende Funktion haben können. Daher müssen eigene hemmende Zellen vorgesehen werden. Wegen der oben angedeuteten Schwierigkeit, geeignete Lernprozesse zu finden, sind diese Zellen immer in einen Kontext fester, nicht durch Lernprozesse zu verstärkender Verbindungen eingebettet. Die hemmende Wirkung ist typischerweise stark.

Sequenzenbildende Zellen. Sequenzen von Erwartungen müssen, wie oben argumentiert worden ist, durch einen eigenen Zelltyp realisiert gedacht werden. Der Lernvorgang entspricht dem der instanzenbildenden Zellen, die Abnahmerate der EPSPs muss aber wesentlich geringer sein als dort, da ja die Zeitspanne zwischen zwei Inputimpulsen überbrückt werden muss, während das für die Kategorisierung gerade nicht wünschenswert ist.

ODER-Zellen. Die Bezeichnung charakterisiert die Funktion. Jeder funktionsfähige Eingang löst ein überschwelliges Aktionspotenzial

aus. Der Lernprozess ist nicht identisch mit dem der instanzenbildenden und der sequenzenbildenden Zellen: Einer der Eingänge muss eine feste Verbindung bilden. Ein über diese oder auch neu verstärkte Verbindungen ausgelöstes Aktionspotenzial öffnet ein kleines Zeitfenster für Lernvorgänge (siehe die Argumentation oben).

Zur Ergänzung und Veranschaulichung folgt hier eine (was die verwendete Architektur angeht stark vereinfachende) Simulation zu den an ODER-Zellen erforderlichen Lernprozessen. Es wird eine funktionsfähige Verbindung aufgebaut und gezeigt, welcher Effekt entsteht, wenn das Zeitfenster für den Lernvorgang nicht mehr erreicht wird.

Simulation:

Lernvorgang an einer ODER-Zelle

In Zeittakt 3 wird ein Erwartungspotenzial in allen sequenzenbildenden Zellen erzeugt.

Ab Zeittakt 10 führt zunächst ein externer Input über die Zelle *i1* zum Feuern der ODER-Zelle *o3*.

Der in Zeittakt 15 anschließende Input über *i2* führt zur Verstärkung der Verbindung *s2—o1*.

Der in Zeittakt 35 relativ spät kommende Input über *i3* führt nicht zu einer sichtbaren Verstärkung der Verbindung *s3—o1*.

Ab Zeittakt 53 wird der gesamte Vorgang wiederholt, die Verstärkung von *s2—o1* führt jetzt zu einer überschwelligeren Verbindung, und *o1* feuert auch aufgrund des Inputs über *i2*. Die Verbindung *s3—o1* bleibt nach wie vor unverändert.

Wenn ODER-Zellen überhaupt eine Funktion haben sollen, muss ihr Output seinerseits Input für höhere Verarbeitungsstrukturen sein. Wie oben in Abschnitt 2.3.4 dargestellt, muss man für die Repräsentation von Verarbeitungshierarchien annehmen, dass der Kortex dafür entsprechend mehrfach in der Senkrechten durchlaufen wird. Das bedeutet, dass der Output der ODER-Zellen direkt oder indirekt wieder Input in einen Zellbestand sein muss, der dieselbe Struktur hat, wie der, aus dem der Output stammt. Die Eingangsstelle dafür können aber nur die instanzenbildenden Zellen sein, gemäß Abbildung 2.4.7–3. (Verglichen mit der an der Biologie orientierten Darstellung in Abschnitt 2.3.4 ist die Hauptverarbeitungsrichtung für Perception und Kategorisierung in den Simulationsmodellen traditionsgemäß von unten nach oben, also nicht, wie wohl biologisch korrekter, von oben nach unten.)

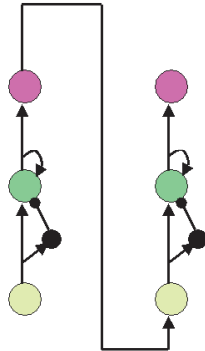


Abbildung 2.4.7–3: Repräsentation von Verarbeitungshierarchien.

2.4.8 Abschließendes Experiment zur Prototypentheorie

Nach den Vorarbeiten in den vorangegangenen Abschnitten kann jetzt mit einer zusammenfassenden Simulationsserie die Frage angegangen werden, wie eine Bewertung eines Exemplars einer prototypischen Kategorie als „besserer Vertreter“ oder „schlechterer Vertreter“ der Kategorie zustande kommt.

Zunächst wird in einer den Vorstellungen des vorangegangenen Abschnitts entsprechenden Architektur der Lernvorgang wiederholt, der oben in Abschnitt 2.4.5 in vereinfachter Form dargestellt worden ist. Der Input entspricht genau dem Input dort.

Architekturdetails: Es gibt 19 zufällig ausgewählte, aber nicht veränderliche Verbindungen von einer Zelle mit der Identifikation st auf sequenzenbildende Zellen, die dort ein Erwartungspotenzial zur Darstellung eines Kontexts erzeugen. Jede der 12 Zellen für Inputmerkmale hat 20 zufällig ausgewählte lernfähige Verbindungen mit instanzbildenden Zellen (Identifikation mit i beginnend). Jede sequenzenbildende Zelle (Identifikation mit s beginnend) hat 30 zufällig ausgewählte lernfähige Verbindungen mit ODER-Zellen (Identifikation mit o beginnend), zusätzlich zu der einen festen überschwelligen Verbindung; außerdem hat jede sequenzenbildende Zelle 5 zufällig ausgewählte lernfähige Verbindungen auf andere sequenzenbildende Zellen. Jede ODER-Zelle hat 5 zufällig ausgewählte lernfähige Verbindungen.

dungen mit instanzenbildenden Zellen. Die übrigen festen Verbindungen entsprechen denen in Abbildung 2.4.7–3. Das heißt, jede instanzenbildende Zelle hat eine knapp unterschwellige Verbindung mit einer zugeordneten sequenzenbildenden Zelle und eine überschwellige Verbindung mit einer hemmenden Zelle. Die hemmenden Zellen haben Verbindungen mit ihren zugeordneten sequenzenbildenden Zellen. Diese Verbindungen sind so stark, dass sie stehengebliebene Abfragepotenziale in den sequenzenbildenden Zellen ausgleichen. Jede sequenzenbildende Zelle hat eine feste unterschwellige Verbindung auf sich selbst zurück.

Simulation:

Lernvorgang KANNE

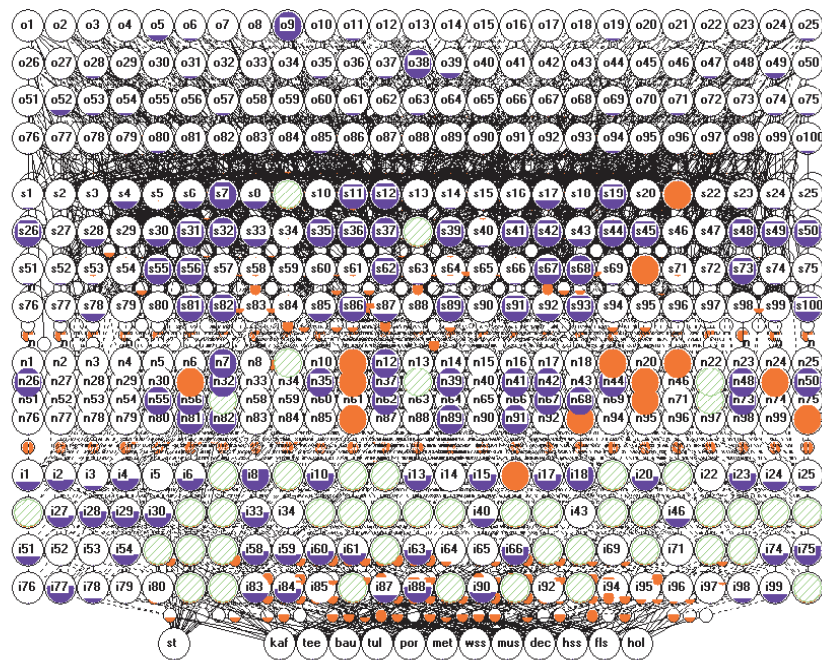
Diese Simulation erfordert eine Bildschirmauflösung von mindestens 1024*768 Pixel.

Ergänzender Hinweis: Steuerung über die Leertaste ergibt den besten Eindruck vom Simulationsverlauf.

Wenn man die Simulation verfolgt, kann man zweierlei feststellen:

- Nicht alle gebildeten Instanzen passen zu dem durch die Erwartungspotenziale auf den sequenzenbildenden Zellen dargestellten Kontext, dieser Kontext wirkt als (vereindeutigender) Filter.
- Die durch einen Input ausgelöste Aktivität läuft mehrfach durch die Zellanordnung, das heißt, es wirkt sich aus, dass die ODER-Zellen Verbindungen zurück auf instanzenbildende Zellen haben.

Ansonsten ist, abhängig von der großen Konnektivität und der Effektivität des Lernvorgangs, die zu erwartende weitgestreute Aktivierung von Zellen zu beobachten. Einen entsprechenden Simulationszustand zeigt die Abbildung 2.4.8–1. Dort zeigen die Zellen *o9* und *o38* ein überschwelliges exzitatorisches Potenzial, das heißt, sie werden im nächsten Zeittakt ein Aktionspotenzial abgeben. Da die Zellen *s21* und *s70* gerade feuern, ist damit zu rechnen, dass einige Zeittakte später auch *o21* und *o70* feuern werden. In dem folgenden Zeittakt 1120 feuern auch noch die in 1119 überschwellig erregten sequenzenbildenden Zellen *s32*, *s37* und *s68*, mit analogen Konsequenzen.



1119

Abbildung 2.4.8–1: Simulationszustand zum Erwerb prototypischer Kategorien.

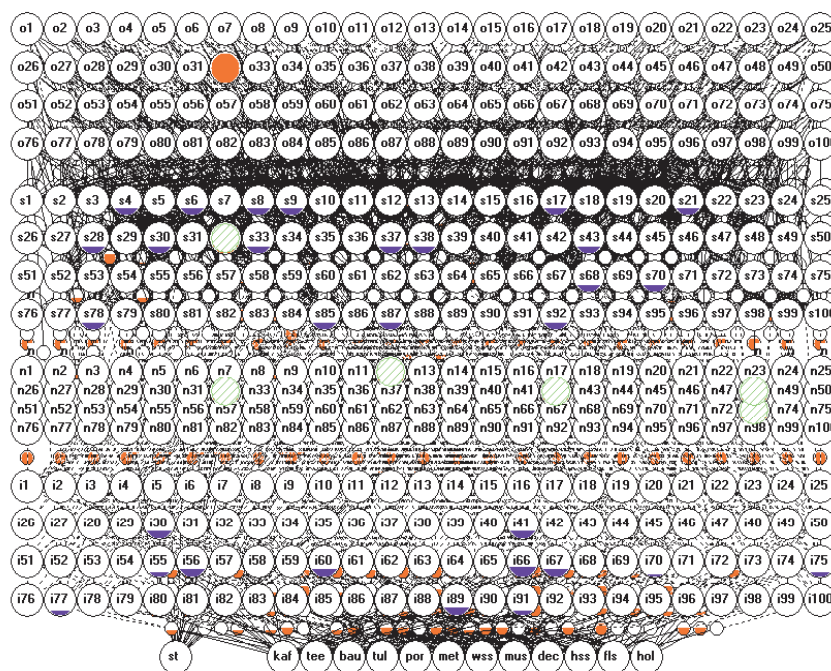
Der besseren Übersichtlichkeit halber wird für die folgenden Simulationen das Ergebnis des Lernvorgangs als Anfangsarchitektur verwendet. Es handelt sich nicht um ein Einfrieren des Lernergebnisses, die Lernparameter sind weiter wirksam. Zunächst soll ein Versuch gezeigt werden mit der Eingabe eines Merkmalsatzes, wie er in der „Lernphase“ verwendet worden ist und der als Repräsentation einer „guten“ Kanne gelten kann (*kaf bau tul por wss mus hss*). Es feuern in diesem Versuch alles in allem 26 ODER-Zellen.

Simulation:
 Reaktion auf gute Kanne
 Diese Simulation erfordert eine Bildschirmauflösung von mindestens 1024*768 Pixel.
 Ergänzender Hinweis: Steuerung über die Leertaste ergibt den besten Eindruck vom Simulationsverlauf.

Als „schlechte“ Kanne kann man ein Exemplar betrachten, dass nur *eines* der für KANNE möglichen Merkmale aktiviert. Mit diesem Input ist nicht

nur das gesamte Aktivationsmuster des Netzes wesentlich ausgedünnt, sondern es wird auch nur noch eine einzige ODER-Zelle zum Feuern gebracht (Abbildung 2.4.8–2, man vergleiche mit Abbildung 2.4.8–1).

Simulation:
 Reaktion auf schlechte Kanne
 Diese Simulation erfordert eine Bildschirmauflösung von mindestens 1024*768 Pixel.
 Ergänzender Hinweis: Steuerung über die Leertaste ergibt den besten Eindruck vom Simulationsverlauf.



18

Abbildung 2.4.8–2: Simulationszustand nach Input eines einzelnen Merkmals.

Man beachte jetzt, dass der Input nicht einfach deshalb eine „schlechte“ Kanne repräsentiert, weil er nur in einem einzigen Merkmal besteht. Es ist für einen Großmutterknoten mit der Bedeutung KANNE gleichgültig, wie viele Eingänge aktiviert sind, wenn er überschwellig erregt wird. Es ist auch unmöglich, sich einen Bewertungsmechanismus vorzustellen, der

überprüft, wie viele Eingänge (ggf. auch überschüssige) nun tatsächlich zum Feuern des Knotens geführt haben und der daraus eine Maßzahl gewinnt. Die Bewertung muss eine andere Basis haben.

Die Lösung ergibt sich aus der einfachen Beobachtung, dass je nach Qualität des Inputs *mehr oder weniger* Großmuttereinheiten aktiviert werden (vgl. dazu auch Kochendörfer, 2000:101). Auch für die Auswertung dieses Effekts kann man allerdings nicht mit einer Art Zählleinrichtung rechnen. Es muss hinzugezogen werden, dass die Zahl der Großmuttereinheiten auch einen Einfluss darauf hat, in wie vielen Kontexten ein dadurch repräsentiertes Konzept vorkommen kann bzw. erlebt worden ist. Es kann dann über Vorstellungsprozesse, wie sie im folgenden Kapitel beschrieben werden, überprüft werden, wie gängig bzw. typisch ein Exemplar ist: Ein sehr gutes Exemplar triggert sehr viele Kontexte, die über Vorstellungsprozesse aufgerufen werden können. Es sind Vergleiche verschiedener Exemplare möglich, die zur Etablierung einer Rangfolge führen können – und damit zu einer Bewertung, wie sie in den Experimenten z. B. bei Rosch (1975) beschrieben wird.

Neben der Bewertung ist oben am Ende von Abschnitt 2.4.5 auch noch der Mechanismus des Erkennens „defekter“ Exemplare ungeklärt geblieben. Die Beobachtung des Zusammenhangs zwischen Kontext und Kategorisierung kann wenigstens einen Teil der Fälle verständlicher machen. Es ist leicht zu zeigen, dass eine Ausweitung des *passenden* Kontexts zu einer Vereinfachung der Kategorisierung führen kann. In dem folgenden Simulationsexperiment wird zunächst ein Kontext erzeugt durch Eingabe des Merkmalsatzes für eine typische Kanne. Anschließend folgt die Eingabe des einen Merkmals aus dem Experiment mit der „schlechten“ Kanne. Während die „schlechte“ Kanne allein nur eine einzige ODER-Zelle überschwellig aktiviert hatte, sind es jetzt drei.

Simulation:

Schlechte Kanne in Kontext

Diese Simulation erfordert eine Bildschirmauflösung von mindestens 1024*768 Pixel.

Ergänzender Hinweis: Steuerung über die Leertaste ergibt den besten Eindruck vom Simulationsverlauf.

In Kochendörfer (2000:102) ist das so interpretiert worden, dass auch unvollständig wahrgenommene Geschirrtteile im Geschirrladen noch als solche kategorisiert werden können, was in einem anderen Kontext nicht möglich wäre.

2.5 Vorstellungen

2.5.1 Produktion von Vorstellungen als Umkehrung der Perzeption

Wenn man von Vorstellungen spricht, bezieht man sich auf produktive Prozesse oder vielleicht noch genauer auf das Ergebnis solcher Prozesse. Vorstellungsprozesse sind, wenn man diese Bestimmung akzeptiert, das Gegenstück von Wahrnehmungsprozessen. Das Ergebnis eines Wahrnehmungsprozesses (oder das, was aufgrund eines Lernvorgangs von einem Wahrnehmungsprozess übrig bleibt) verdient nur dann die Bezeichnung „Vorstellung“, wenn es produktiv, also losgelöst von dem Wahrnehmungsinput, aktiviert wird, z. B. in einem Vorgang des Sich-Erinnerns.

Nicht alle Vorgänge, die als Umkehrung von Wahrnehmungsprozessen gelten müssen, werden in der Regel als „Vorstellungen“ akzeptiert. So ist selbstverständlich die Sprachproduktion eine Spiegelung von Perzeptionsprozessen, auch dann, wenn es sich nicht um eine wörtliche Wiedergabe von sprachlichen Wahrnehmungen handelt, man denke an die Funktion und Wirkung von Spracherwerbsprozessen als Grundlage aktueller Produktion. Sprechen wird allerdings in der Regel als motorische Produktion gesehen und ist insofern eine Form „äußeren“ Handelns, während Vorstellungen als „innere“ Prozesse gelten. Eine Abgrenzungsschwierigkeit wird aber deutlich, wenn man an das sogenannte „innere Sprechen“ denkt, das als so etwas wie die Vorstellung des Sprechens aufgefasst werden kann, oder an das „rehearsal“ in Modellen des Kurzzeitgedächtnisses (vgl. Kochendörfer, 1999; im Zusammenhang mit Vorstellungen: Smith, Wilson & Reisberg, 1995).

Wahrscheinlich wird auch bei Formulierungen wie der von Kosslyn (1995: 267) nicht in erster Linie an Sprache gedacht:

„Because most research has focused on visual imagery, we will concentrate on that type of imagery here, although many people report

experiencing imagery in all sensory modalities (“hearing” with the “mind’s ear,” “tasting” with the “mind’s tongue,” and so on).“

Trotzdem wird man sprachliche Vorstellungen, wie andere akustische Vorstellungen auch, mit in die Betrachtung einbeziehen.

Gerade das Beispiel des inneren Sprechens, das nicht einfach Wiederholung von Auswendiggelerntem ist, kann als Hinweis darauf dienen, dass Vorstellungen grundsätzlich, das heißt auch im außersprachlichen Bereich, nicht immer eine Wiederbelebung von vollständig schon vorhandenen Gedächtnisinhalten darstellen müssen, sondern prinzipiell, wahrscheinlich sogar in den meisten Fällen, auch durch *konstruktive* Prozesse entstehen. „Konstruktiv“ meint z. B., dass kleinere Elemente, die Gedächtnisinhalte (Ergebnisse von Lernprozessen) sind, im Vorstellungsprozess in einen neuen, größeren Zusammenhang gebracht werden.

Sinnesprimitive, die ja schon definitionsgemäß auf die Seite der Perzeption gehören, können nicht Elemente von Vorstellungen sein, es braucht mindestens eine interne Repräsentation, die den Sinnesprimitiven entspricht: Vorstellungen entstehen nicht durch eine Anregung der Netzhaut, oder im Fall akustischer Vorstellungen durch eine Anregung der Sinneszellen des Innenohrs. Zu diskutieren ist die Art der internen Repräsentation und die Art des Zustandekommens der Vorstellung selbst. Wenn man zunächst davon ausgeht, dass Vorstellungen vollständig dadurch zu beschreiben sind, dass sie in der Aktivierung einer internen Repräsentation bestehen, muss man sich mit zwei Möglichkeiten auseinandersetzen. Kosslyn (1995: 280):

„Two means of representation have been proposed for mental images, one that confers a special status on images and one that treats them as no different in kind from the representations of linguistic meaning. The two alternatives are called *depictive* and *propositional* representation [...].“

Kosslyn argumentiert schlüssig für die „depictive representation“, indem er zunächst auf mehrere neurowissenschaftliche Befunde zurückgreift:

„First, it has long been known that some visual areas of the brain are topographically organized. These regions of cortex preserve the spatial structure (roughly) of the retina [...]. Second, it has been found that connections between visual areas typically do not simply send information downstream. Rather, these connections usually run in both directions. Third, the areas of the brain that store visual memories are not topographically organized.“ (Kosslyn, 1995: 290)

Auf einer der Retina näheren Ebene der Verarbeitung gilt also das retinotopische Prinzip, auf höheren Verarbeitungsebenen („visual memory“) ist das nicht mehr der Fall. Es gibt Verbindungen nicht nur von der primären Verarbeitung zum visuellen Gedächtnis, sondern auch in der Gegenrichtung.

Die entscheidenden Beobachtungen stammen dann aus PET-Experimenten, die zeigen, dass bei visuellen Vorstellungsaufgaben und bei geschlossenen Augen die topographisch organisierten Kortexbereiche, also nicht nur Kortexbereiche, für die man eine propositionale Repräsentation annehmen könnte, aktiviert sind.

„Moreover, the precise regions of activation were quite close to what one would expect based on the results from visual stimulation studies in the literature.“ (Kosslyn, 1995: 291)

Ergänzend ist gezeigt worden, dass durch eine Schädigung der betreffenden Regionen Vorstellungen beeinträchtigt werden. (Vgl. insgesamt auch Kosslyn, 1994.)

Es gibt einige Arbeiten, die eine Gegenposition vertreten bzw. die experimentellen Ergebnisse nicht bestätigen. Wie immer ist dabei darauf zu achten, dass wirklich die für die Theorie relevanten Bedingungen eingehalten werden. Das ist z. B. bei Knauff et al. (2000) nicht der Fall und bei Trojano et al. (2000) fraglich.

Offen bleibt zunächst, ob auf der höheren Ebene des „visual memory“ nicht doch noch propositionale oder jedenfalls nicht-depiktive Repräsentationen eine Rolle spielen. Kosslyn (1995: 292):

„It is clear, that the depictive representations are one component of the ensemble of representations and events that underlie imagery, but they are not the only component.“

Wenn die Produktion von Vorstellungen in irgendeiner Weise als Spiegel der Perception von (z. B. visuellen) Wahrnehmungen gesehen werden kann, und wenn Kategorien Eigenschaften haben, wie in den vorangegangenen Kapiteln besprochen, ist mit einer propositionalen Repräsentation nicht notwendig zu rechnen. Es ist auch nicht klar, wie propositionale Repräsentationen neuronal darzustellen sind. Man sollte sich nicht täuschen: Wenn man sagen kann „die Pyramide steht auf dem Würfel“ dann muss das nicht den Besitz einer Proposition wie (STEHEN-AUF, PYRAMIDE, WÜRFEL) bedeuten. Einfacher ist es, mit einer Großmutterzelle KLOTZWELT zu rechnen, die in der visuellen Wahrnehmung einer entsprechenden Anordnung aktiviert wird durch gleichzeitige Aktivität von Großmutterzellen für WÜRFEL-UNTEN

und PYRAMIDE-OBEN. Ein weiteres Argument, das in dieselbe Richtung führt, ergibt sich aus der Überlegung, dass auch akustische Wortformen, also Lautketten, vorgestellt werden können. Diese Wortformen müssen im sprachlichen Lexikon gespeichert vorliegen, niemand wird aber annehmen, dass das durch eine Folge von so vielen Propositionen geschieht, wie eine solche Wortform lautliche Segmente enthält.

2.5.2 Rückspiegelung

Die Aktivierung (was immer man sich darunter vorstellen mag) einer Großmutterzelle ergibt noch kein Bild. Bildhafte Vorstellungen brauchen Informationen, die in der einzelnen Zelle nicht repräsentiert sind. Es muss sich um Informationen handeln, die, in Perzeptionsrichtung gesehen, zur Definition der Großmutterzelle beitragen bzw. beigetragen haben. Andererseits muss eine spezifische Vorstellung, sofern es sich um eine bereits etablierte Ganzheit handelt, von einer spezifischen Großmutterzelle *ausgehen*, die diese Vorstellung repräsentiert. Es müssen also Vorgänge angenommen werden, die von einer solchen Großmutterzelle nach außen führen, in Richtung auf die Sinnesorgane zu. Diese Vorgänge setzen entsprechende neuronale Verbindungen voraus. Weil neuronale Verbindungen immer unidirektional sind, muss ein eigenes Bahnensystem (einschließlich der dafür erforderlichen Neuronen) für die Produktion vorhanden sein, das zu den durch Lernen verstärkten Perzeptionsbahnen parallel verläuft.

Man könnte nun vielleicht annehmen, dass eine Vorstellung darin besteht, dass eine spezifische Produktionsbahn (ggf. mit entsprechend zahlreichen Verzweigungen) durch eine in Produktionsrichtung verlaufende Impulsfolge aktiviert wird, die sich ggf. wiederholt. Die Produktionsbahn wäre damit von der neuronalen Umgebung abgehoben und der Vorstellungsvorgang bzw. die neuronale Repräsentation einer Vorstellung würde sich darin erschöpfen. Es muss jetzt aber ergänzend beachtet werden, dass aus Vorstellungen auch Schlussfolgerungen gezogen werden können, Vorstellungen sollen zu Konsequenzen führen können, die ein Stück weit den Konsequenzen „echter“ Wahrnehmungen entsprechen. Das ist z. B. in Vorgängen der Fall, die als „Probehandeln“ bezeichnet werden. Solche Schlussfolgerungen müssen durch Lernvorgänge vorbereitet sein, das heißt, es müssen entsprechende neuronale Verbindungen vorhanden sein, die sie ermöglichen. Ausgangspunkt solcher Verbindungen müssen aber Perzeptionsstrukturen sein, nicht Produktionsstrukturen, sonst müsste man eine doppelte Verankerung von Erfahrungen annehmen.

Man kann die Wirkung von Vorstellungen offenbar nur dann erklären, wenn man eine *Rückspiegelung* aus der Produktion in die Perzeption annimmt (Abbildung 2.5.2–1). Visuelle Vorstellungen werden dadurch bildhaft, dass sie buchstäblich in die visuelle Verarbeitung einmünden oder besser: sie *entstehen* eigentlich erst dadurch, dass durch Produktionsvorgänge ein Ersatz für eine äußere visuelle Erregung geschaffen wird. Es ist die Stärke der im vorigen Abschnitt behandelten Beobachtungen von Kosslyn, dass sie dieser aus den neuronalen Verarbeitungsbedingungen geschlossenen Rückspiegelung genau entsprechen.

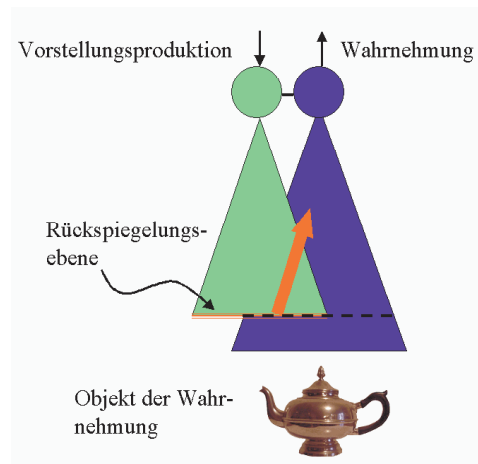


Abbildung 2.5.2–1: Schema zur Bildung visueller Vorstellungen durch Rückspiegelung in die Wahrnehmungsbahnen. Der Kegel der Bahnen, die zur Definition einer Großmutterzelle in der Perzeption führen und die parallelen, von einer „Kopie“ dieser Großmutterzelle ausgehenden Produktionsbahnen sind durch Dreiecke dargestellt. Die Rückspiegelung der Produktionssignale geschieht auf einer Ebene vor Erreichen der Sinnesorgane.

Die Rückspiegelung der Ergebnisse von Produktionsvorgängen in die Perzeption ist für die Sprache eine z. B. bei Levelt (1989) zentrale Idee. Sie dient dort dem sog. „Monitoring“, das eine wichtige Rolle bei Formulierungsprozessen spielt (Abbildung 2.5.2–2). Bei Levelt gibt es zwei Wege, auf denen etwas „zurückgespiegelt“ werden kann: den Weg über das Ohr und den Weg ausgehend von einer Vorstufe der Artikulation, dem „phonetic plan“. Eine Möglichkeit der Rückspiegelung, die von dieser Vorstufe ausgeht, muss angenommen werden, weil man auf ein Monitoring auch beim inneren (also lautlosen) Sprechen nicht verzichten kann. Oft ist inneres Sprechen mit einer

teilweisen Innervierung der beim lauten Sprechen beteiligten Muskeln verbunden. Das ist aber nicht notwendig der Fall. Wenn man zulässt, dass der innere Weg auch ohne die an die motorische Peripherie heranreichende Aktivierung funktioniert, entspricht er genau dem Rückspiegelungsvorgang bei visuellen Vorstellungen: Visuelle Vorstellungen haben keine Wirkung „nach außen“, daher ist ein Weg für die Rückspiegelung über die Augen oder über die Wahrnehmung von motorischen Phänomenen nicht möglich.

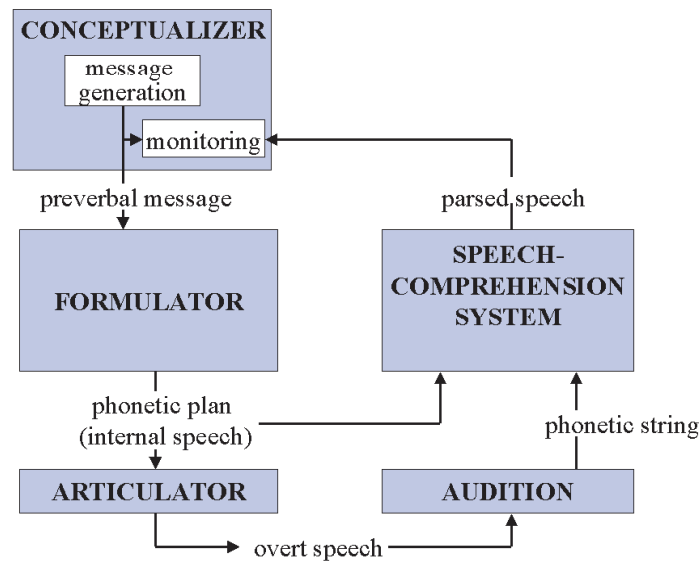


Abbildung 2.5.2-2: Rückspiegelung von internal (und overt) speech für Monitoring-Zwecke in dem Sprachproduktionsmodell von Levelt (1989: 9). Die Darstellung ist auf die in diesem Zusammenhang wesentlichen Komponenten verkürzt.

Vorstellungen erreichen nicht dieselbe Detailliertheit und „Farbigkeit“ wie „echte“ Wahrnehmungen, gleichgültig in welcher Sinnesmodalität sie gebildet werden. Vorstellungen können – um im Bild zu bleiben – nicht „farbiger“ sein als die ihnen zugrundeliegenden Gedächtnisspuren. Diese Spuren sind aber Abstraktionen, deren Detailliertheit von verschiedenen Faktoren abhängig ist, z. B. von der Häufigkeit des Vorkommens einer Wahrnehmung oder von den Intensität einer Wahrnehmung, abhängig von ihrer Bedeutung für das betroffene Individuum. Vor allem visuelle Vorstellungen werden teils als mehr, teils als weniger lebhaft empfunden.

2.5.3 Neuronale Grundlagen und Lernvorgänge

Die für Vorstellungsprozesse erforderlichen neuronalen Strukturen umfassen, wie im vorigen Abschnitt schon angedeutet, zwei Komponenten:

- ein System von Top-down-Verbindungen, das durch Lernprozesse entstehen muss, um der Spezifität der einzelnen Vorstellungen Rechnung zu tragen;
- Verbindungen, die der Rückspiegelung von Aktivitäten dieses Systems in die Wahrnehmungsstrukturen dienen.

Top-down-Verbindungen

Es sollen hier nun zunächst die Top-down-Verbindungen behandelt werden. Zu den diese Verbindungen betreffenden Gesichtspunkten gehört, dass sie unmittelbar in Zusammenhang mit Wahrnehmungsprozessen sozusagen „automatisch“ entstehen müssen. Vorstellungen können nicht erst dann gebildet werden, wenn ein zusätzlicher, spürbarer(!) Übungsprozess stattgefunden hat. Die Top-down-Verbindungen müssen damit als Pendant der entsprechenden Bottom-up-Verbindungen und im Zug der Aktivierung dieser Verbindungen entstehen.

Da die Zellkombinationen, die Perzeptionsvorgänge gewährleisten (vgl. z. B. oben Abbildung 2.4.7–1), zwar in sich fest strukturiert sind, aber spezifische Verbindungen untereinander eingehen, muss, um diese spezifische Verschaltung in Produktionsrichtung abzubilden, pro Zellkombination mindestens eine top-down orientierte Zelle vorhanden sein. Diese Zellen werden im Folgenden als P-Zellen (früher „prädiktive Zellen“, nach einer ihrer Funktionen) bezeichnet. Das Muster der Verbindungen in Perzeptionsrichtung kann als eine Mischung von Konvergenz und Divergenz beschrieben werden. Divergenz von bottom-up gerichteten Bahnen muss in Top-down-Richtung durch Konvergenz an den P-Zellen gespiegelt werden, Konvergenz umgekehrt durch eine von den P-Zellen ausgehende Divergenz, wie in Abbildung 2.5.3–1 schematisch dargestellt.

Wie für die Perzeptionsrichtung müssen auch für die Produktionsrichtung entsprechende potenzielle Bahnenstrukturen vorausgesetzt werden, die einen Lernprozess erst ermöglichen. Mit einer gewissen angeborenen Spiegelbildlichkeit der potenziellen Bahnen für Produktion und Perzeption kann gerechnet werden, beide Bahnsysteme werden aber durch Lernprozesse, also Verstärkung ausgewählter Bahnen, weiter spezifiziert.

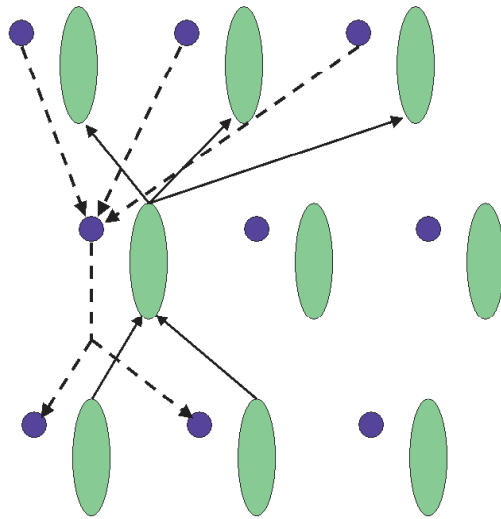


Abbildung 2.5.3-1: Entsprechungen von Konvergenz und Divergenz in Perzeptionsrichtung (Zellkombinationen, die Abbildung 2.4.7-2 entsprechen, sind durch Ovale wiedergegeben) und Produktionsrichtung.

Die genauere Bestimmung der Lernprozesse für das Top-down-System setzt relativ komplexe Überlegungen voraus. Man muss sich zunächst um die Natur der Konvergenzen auf diesem System kümmern. Konvergenz wird im Kortex immer durch synaptische Verbindungen realisiert, die auf einer die Konvergenz gewährleistenden Zelle liegen. Es gibt, wenn man die Art des Funktionierens dieser Zelle betrachtet, verschiedene Formen von Konvergenz:

- (1) Konvergenz, die durch Synapsen entsteht, deren Erregungen gleichzeitig (das heißt innerhalb eines kleinen Zeitfensters) erscheinen und aufsummiert werden, um zum Feuern der Zelle zu führen (wie in Perzeptionsrichtung bei der Konzeptbildung), und
- (2) Konvergenz, bei der eine alternative Erregung verschiedener Eingänge zu einem Feuern der Zelle führt (wie in Perzeptionsrichtung bei der Zusammenfassung von Elementen zu einer Szene, ODER-Verknüpfung).

Divergente Bahnen werden immer als Axonkollateralen gebildet, was zur Folge hat, dass verschiedene divergente Verarbeitungsbahnen in der Perzeption immer gleichzeitig erregt werden. Andererseits gilt, dass natürlich

nicht alle diese Erregungen in nachfolgenden Verrechnungsprozessen erfolgreich bleiben. Man muss deshalb für die Produktion damit rechnen, dass nicht alle konvergenten Produktionsbahnen, die divergenten Perzeptionsbahnen entsprechen, gleichzeitig aktiviert sind, und es ist auch von Fall zu Fall verschieden, welche es sind. Damit kann eine entsprechende Konvergenz in der Produktion nur der Möglichkeit (2) entsprechen: Die produktionsorientierten konvergenten Entsprechungen der verschiedenen perzeptionsorientierten divergenten Bahnen müssen *gleichberechtigt* eine produktionsorientierte Wirkung auslösen können. Die in Abbildung 2.5.3-1 eingezeichneten top-down orientierten Zellen müssen damit ODER-Zellen sein, das heißt Zellen mit dem für ODER-Zellen spezifischen Lernmechanismus.

Diese Entscheidung hat weitere wichtige Konsequenzen. Es ist nicht möglich, ODER-Verbindungen aufzubauen, ohne dass die Zielzelle mindestens *eine* vor dem Lernvorgang funktionsfähige Verbindung hat, deren Erregung zum Feuern der Zelle führt. Dieses Feuern der Zelle bewirkt die Auswahl der Zelle für den Lernvorgang. Die Auswahl muss zur Parallelität der Produktionsstrukturen mit den Perzeptionsstrukturen führen, kann also nur im Zuge eines Perzeptionsvorgangs, bzw. eines Vorgangs, bei dem die Perzeptionsstrukturen aktiviert sind, geschehen, womit klar wird, dass eine Verbindung mit einer Zelle der Perzeptionsstruktur vorausgesetzt ist. Diese Zelle kann prinzipiell irgendeine der Zellen dieser Struktur sein. Für die sequenzenbildende Zelle als Partner spricht die Überlegung, dass die Rückspiegelung eines Produktionsvorgangs nur dann funktionieren kann, wenn die dafür erforderlichen sequenzenbildenden Zellen ein Erwartungspotenzial haben, das der Produktionsvorgang zuvor aufgebaut hat (Kontextfilter!). Also werden unerschwellige Verbindungen von den P-Zellen zu den sequenzenbildenden Zellen vorausgesetzt und man wird am ehesten annehmen, dass es Verbindungen sind, die parallel dazu von den sequenzenbildenden Zellen zu den P-Zellen führen, die als Voraussetzung für den Aufbau von ODER-Verbindungen auf den P-Zellen erforderlich sind. (Diese Überlegung gilt allerdings nur für den Fall, dass ein Kontext für die Perzeption nicht aus einem vorangegangenen Perzeptionsprozess entsteht. Ein stützendes Argument ergibt sich aber auch aus der „prädiktiven“ Funktion der P-Zellen, die in Kochendörfer, 2000: Kapitel 5, beschrieben ist; vgl. auch Teil 5, „Syntax“.)

Wie auch immer man sich entscheidet, gilt für diese Verbindungen in jedem Fall, dass sie unerschwellig sein müssen und allein nicht zum Feuern der P-Zellen führen dürfen, weil ja nicht jede Perzeption zu einer gleichzeitigen Produktion führen soll. Sie können selbstverständlich auch nicht durch Lernprozesse entstehen, wenn der Lernprozess an den Zielzellen (den

P-Zellen) dem ODER-Prozess entspricht: Sie sollen ja einen solchen Lernprozess allererst ermöglichen. Auch die Verbindungen von den P-Zellen zu den sequenzenbildenden Zellen können nicht über Lernprozesse verstärkt werden, weil der Zeitpunkt des Feuerns der P-Zellen nach dem Feuern der sequenzenbildenden Zellen liegt, und der für diese Zellen gültige Lernprozess die umgekehrte Reihenfolge voraussetzen würde. Wir gewinnen damit die Annahme, dass eine bidirektionale (das heißt natürlich: über zwei gegenläufige Bahnen hergestellte) feste Verbindung zwischen P-Zellen und sequenzenbildenden Zellen existieren muss.

Die Unterschwelligkeit der festen Verbindung von den sequenzenbildenden Zellen zu den P-Zellen hat zur Folge, dass P-Zellen schon einen gewissen erregenden Potenzialpegel haben müssen, wenn sie aufgrund der Erregung dieser Verbindung feuern sollen, um einen Lernprozess zu ermöglichen. Man muss sich überlegen, woher dieser Potenzialpegel kommen kann. Es sind zwei Quellen denkbar: Die einfachste Möglichkeit ist, dass die entsprechende Verbindung selbst diesen Potenzialpegel aufbaut, dadurch, dass sie ausreichend kurz hintereinander zwei Aktionspotenziale führt. Diese Möglichkeit wird im Folgenden als „Doppelwellenprinzip“ bezeichnet. Die zweite Möglichkeit ist die, dass man den Pegel durch ein weiteres System erzeugen lässt, das vorläufig als „Hilfssystem“ bezeichnet wird.

Der damit konstruierbare Lernprozess kann in den vereinfachten Simulationsarchitekturen der Abbildung 2.5.3–2 demonstriert werden. Hier sind von den für die Perzeption erforderlichen Zellen nur die sequenzenbildenden Zellen aufgenommen, Verbindungen zwischen diesen Zellen gelten als bereits etabliert. Der simulierte Lernprozess betrifft die Top-down-Verbindungen zwischen den P-Zellen, die in der Abbildung gestrichelt eingezeichnet sind.

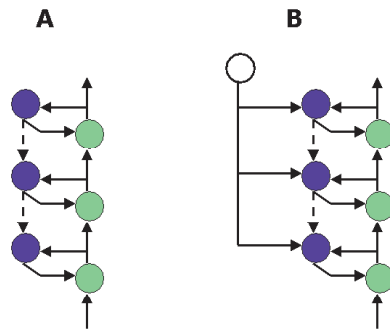


Abbildung 2.5.3–2: Beispielarchitekturen zur Simulation des Erwerbs von Top-down-Verbindungen. **A:** nach dem Doppelwellen-Prinzip. **B:** mit Hilfssystem.

Simulationen:
 Simulation nach dem Doppelwellen-Prinzip
 Simulation mit Hilfssystem

Ergänzender Hinweis:
 Auf den Simulationsbildschirmen ist wegen der Überlagerung der Symbole nur eine der Verbindungen zwischen P-Zellen und sequenzenbildenden Zellen gezeigt, die Simulation enthält in Wirklichkeit beide. Steuerung über die Leertaste ergibt den besten Eindruck vom Simulationsverlauf.

Die Existenz von Top-down-Verbindungen, wie sie durch den beschriebenen Lernprozess entstehen, wäre ausreichend zur Gewährleistung von Vorstellungen, wenn bei einer Verzweigung in Top-down-Richtung immer alle Zweige zu aktivieren wären. Das ist aber nicht immer der Fall. Am Beispiel einer Szene: Es ist nicht sinnvoll, alle Teile einer Szene gleichzeitig zu aktivieren. Es sind Strukturen erforderlich, die es erlauben, spezifische Top-down-Bahnen auszuwählen. Die Auswahl von Bahnen geschieht in Perzeptionsrichtung über den Kontextfilter, das heißt über den Mechanismus der sequenzenbildenden Zellen. Eine Entsprechung dazu muss nun offenbar auch in Produktionsrichtung vorhanden sein.

Dabei könnte man zunächst an Kontextverbindungen direkt auf den P-Zellen denken. Kontextverbindungen auf den P-Zellen als ODER-Zelle sind aber nicht durch Lernprozesse zu etablieren, da die kontextetablierenden Vorgänge dem Feuern der P-Zelle vorangehen, nicht folgen. Also muss für die Kontextauswertung ein eigener Zelltyp eingeführt werden, dessen Lernprinzip den sequenzenbildenden Zellen der Perzeption entspricht, der aber Teil des Produktionsmechanismus ist. Zellen dieses Typs werden im Folgenden als M-Zellen bezeichnet.

Parallel zu den Überlegungen bei der Kontextsteuerung in der Perzeption (vgl. oben 2.4.6) gilt auch für die Produktion, dass ein Hemmmechanismus vorgesehen werden muss, der das „Stehenbleiben“ des Kontext-Abfragesignals in den M-Zellen verhindert. Das Kontext-Abfragesignal ist für die Perzeption inputgesteuert, das heißt, es sind nur solche Zellen betroffen, die dem gerade bearbeiteten Input entsprechen. Für die Produktion muss es inhaltsneutral, das heißt zentral ausgelöst und in größeren Bereichen global wirksam sein. Es ist in früheren Publikationen als „Produktionstakt“ bezeichnet worden. M-Zellen können jeweils entsprechenden P-Zellen fest zugeordnet werden. Die Verbindung von der jeweiligen M-Zelle zur P-Zelle muss unterschwellig sein. Das ergibt die Konstruktion der Abbildung 2.5.3-3. (Diese Konstruktion ist schon in Kochendörfer, 1997: 277 f., mit einer anderen Begründung für lexikalische Ausdrucksseiten eingeführt worden.)

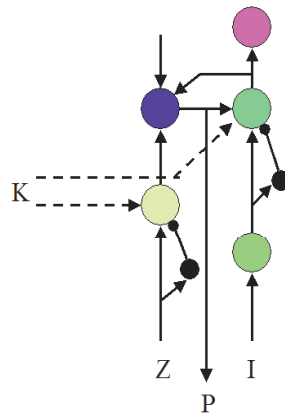


Abbildung 2.5.3–3: Architektur mit Kontextkontrolle im Produktionsbereich.
 K = Kontextverbindungen, Z = zentrale Auslösung, P = Verbindung mit Produktionsprozeduren (Motorik), I = Perzeptionsinput.

Zusätzlich muss jetzt beachtet werden, dass die Kontextverbindungen im Produktionsbereich parallel zu den Kontextverbindungen im Perzeptionsbereich und zusammen mit ihnen entstehen müssen.

Kontextverbindungen, auch die auf den M-Zellen, werden im Zug der Perception aufgebaut, es wird dabei vorausgesetzt, dass die M-Zellen feuern; das tun sie nur, wenn die zentrale Auslösung erfolgt. Das wiederum hat zur Folge, dass über die Verbindung mit den P-Zellen eine unterschwellige Erregung auf diesen Zellen entsteht, die dann ggf. genau die Wirkung hat, die in dem Simulationsversuch oben durch das Hilffsystem gewährleistet worden ist. Es liegt nahe, das Hilffsystem mit dem Produktionstaktsystem zu identifizieren. Der Aufbau des Top-down-Systems der P-Zellen kann dann dadurch bewirkt werden, dass das Produktionstaktsystem zentral aktiviert wird.

Damit ist nicht impliziert, dass der Aufbau des Top-down-Systems der P-Zellen grundsätzlich an eine Aktivität des Produktionstaktsystems gebunden ist. Es bleibt die Möglichkeit des Aufbaus nach dem Doppelwellenprinzip. Entsprechende Aktivitäten werden in den Simulationsmodellen in solchen Fällen beobachtet, wo auf einer höheren Verarbeitungsebene in der Perception mehrere Inputaktivitäten leicht zeitlich versetzt über ODER-Zellen konvergieren. Solche Phänomene müssen dann zwangsweise zum Aufbau paralleler Produktionsstrukturen führen. Man muss also nicht notwendig entscheiden, ob prinzipiell die „Hilffsystem“-Methode oder das „Doppelwellenprinzip“ für den Aufbau von Produktionsstrukturen gilt.

Eine etwas genauere Analyse der Struktur von Abbildung 2.5.3–3 ergibt, dass in Produktions- und in Perzeptionsrichtung gleichartige Teilstrukturen identifiziert werden können, die jeweils eine ODER-Zelle, eine sequenzenbildende Zelle und eine hemmende Zelle enthalten. Das ist bemerkenswert, wenn man die Meinung vertritt, dass Produktion und Perzeption – vielleicht auch nur teilweise – an verschiedenen Stellen des Kortex angesiedelt sind, weil die anatomischen Strukturen des Kortex in allen Bereichen relativ vergleichbar sind. Unterschiede zwischen der Perzeptions- und der Produktionsseite betreffen z. B. das Detail der hemmenden Verbindungen. Auf der Produktionsseite müssen hemmende Zellen und sequenzenbildende Zellen nicht eins-zu-eins zugeordnet sein, sondern hemmende Zellen können Verbindungen auf mehreren sequenzenbildenden Zellen bilden.

Eine unterschiedliche Lokalisierung von Perzeptions- und Produktionsstrukturen ist nicht undenkbar, man beachte, dass nur wenige Verbindungen zwischen Perzeption und Produktion erforderlich sind, deren weitreichende Teile angeboren sein können. (Nur die Sequenzverbindung unterliegt überhaupt Lernprozessen; siehe Abbildung 2.5.3–4.)

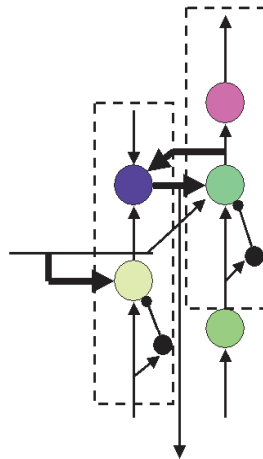


Abbildung 2.5.3–4: Identifizierung gleichartiger Teilstrukturen in Perzeption und Produktion. Verbindungen, die entsprechend weitreichend gedacht werden müssen, wenn Perzeption und Produktion an unterschiedlichen Stellen angesiedelt sind, sind fett ausgezeichnet.

Rückspiegelungsverbindungen

Die Struktur und das Zustandekommen der Rückspiegelungsverbindungen läßt sich wenigstens für den Bereich des inneren Sprechens etwas genauer spezifizieren, wenn man die folgenden Gesichtspunkte beachtet:

1. Es ist nur *eine* Rückspiegelungsebene anzunehmen; diese Rückspiegelungsebene muss relativ weit „außen“ liegen. Wenn man mehrere Rückspiegelungseingänge annimmt, die schematisch an P-Zellen gebunden sind, wo immer eine solche Anbindung möglich ist, entstehen sehr hochfrequente, an Epilepsien erinnernde Erregungen. Das lässt sich leicht in der Simulation zeigen und gilt nicht nur für das innere Sprechen. Die in Abbildung 2.5.3–5 wiedergegebene Simulationsarchitektur ist insofern vereinfacht, als M-Zellen weggelassen sind. Die erforderlichen Erwartungspegel werden durch die nicht eingefärbte Zelle erzeugt, die auf dem Simulationsbildschirm nicht sichtbar ist. Der Start eines Produktionsprozesses wird durch externen Input auf die oberste der P-Zellen bewirkt. Rückspiegelungsverbindungen sind fett ausgezeichnet. Bei der Simulation der Rückspiegelung über eine einzelne Ebene sind die oberen beiden Rückspiegelungsverbindungen weggelassen.

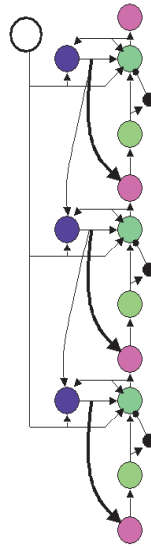


Abbildung 2.5.3–5: Architektur zur Simulation von Rückspiegelungsvorgängen. Erläuterungen im Text.

Simulationen:
 Rückspiegelung über mehrere Ebenen
 Rückspiegelung über eine einzelne Ebene
 Ergänzender Hinweis: Steuerung über die Leertaste ergibt den besten Eindruck vom Simulationsverlauf.

2. Da Produktions- und Perzeptionsbahnen sich im sprachlichen Bereich unterhalb der Phonemebene räumlich trennen und auch strukturell abweichen (Abbildung 2.5.3–6), kann es unterhalb dieser Ebene keine direkte, durch Lernvorgänge ermöglichte Einwirkung der Produktion auf die Perzeption geben. Letzte gemeinsame Station mit direkter Äquivalenz von Bottom-up- und Top-down-Prozessen ist die Phonemebene. Argumente, auf denen diese Annahme beruht, sind: die Einzelsprachlichkeit der Phoneme (Phoneme sind nicht angeboren, müssen also in Strukturen, für die Einzelimpulskodierung gilt, gelernt werden), die Unterschiedlichkeit der phonetischen Merkmale für Perzeption und Produktion, und die Vorgänge (und deren Begründung) in der Lallphase des Spracherwerbs (vgl. Kochendörfer, 2002: 79 f.).

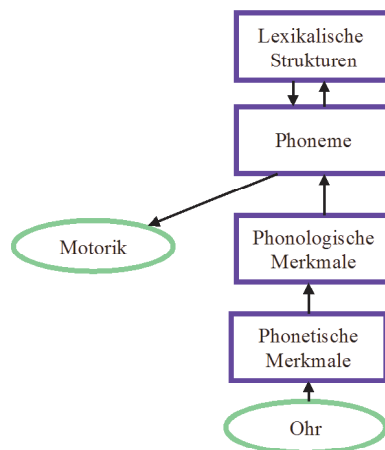


Abbildung 2.5.3–6: Trennung von Produktions- und Perzeptionsbahnen in peripheren Bereichen der Sprachverarbeitung.

3. Man könnte daran denken, dass für die Rückspiegelung direkt die Rückmeldung aus den Sinnesorganen der Muskelfasern verantwortlich ist. Diese Rückmeldung ist aber nicht sprachspezifisch, das heißt, es gibt eine schwer

durch Lernprozesse zu überbrückende Verbindungslücke von den nichtsprachlichen Sinneswahrnehmungen zur (sprachspezifischen) Phonemebene. Details sind der Abbildung (2.5.3-7) zu entnehmen, die Verbindungslücke ist dort durch ein Fragezeichen angedeutet.

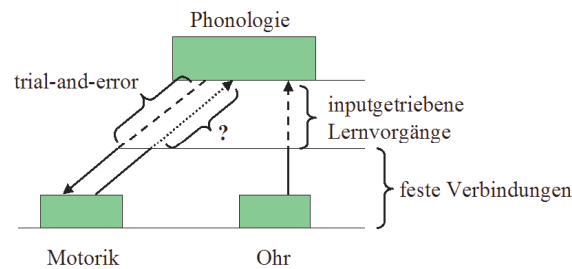


Abbildung 2.5.3-7: Erklärungslücke bei der Annahme einer Rückspiegelung über die Sinnesorgane der Muskelfasern.

Im (normalen) Spracherwerb erfolgt die Kontrolle der Korrektheit der Artikulation in dem in der Lallphase greifbaren Trial-and-Error-Prozess über das Ohr und den Vergleich mit Wahrnehmungen der Umgebungssprache, deren neuronale Grundlage jeweils bereits etabliert ist. Eine sprachspezifische Strecke der Körperwahrnehmung kann aber nicht vor dem Zeitpunkt der Artikulation verfügbar sein. Sie kann bestenfalls zu Stande kommen, nachdem korrekte Artikulationsprogramme bereits entwickelt sind, und muss zu Verbindungen mit auditiven Perzeptionsbahnen führen, die ebenfalls bereits vorhanden sind. Wenn man realistischerweise an mehrere neuronale Zwischenstufen denkt, ist ein Trial-and-Error-Prozess für den Aufbau der Rückmeldungsbahnen erforderlich, für den unklar bleibt, worin Erfolg und Misserfolg bestehen.

Es ist nicht zu bestreiten, dass es bei innerem Sprechen zu Aktivierungen der Motorik kommen kann. Man hat beim inneren Sprechen oft eher den Eindruck eines motorischen Ereignisses als den eines Schallereignisses. Die daraus resultierenden Rückmeldungen können aber nur begleitende Funktion haben, das heißt, sie gehören zu einem begleitenden, nicht sprachspezifischen motorischen Wahrnehmungsbild.

4. Die Rückspiegelung findet an einer Stelle statt, an der es vor einem Lernprozess, der die entsprechenden Rückspiegelungsverbindungen betrifft, bereits etablierte (feste) Verbindungen zur Sinnesperipherie gibt. Die Entwicklung der Perzeption geht in jedem Fall der Rückspiegelung voraus. Wenn die Rückspiegelungsverbindungen auf instanzbildende Zellen führen würden,

dürften diese Zellen vor dem Lernvorgang für Rückspiegelungsverbindungen nicht gefeuert haben, was aber notwendigerweise der Fall sein muss. Dieses Argument schließt auch sequenzenbildende Zellen in den Perzeptionsbahnen als Ziel von Rückspiegelungsverbindungen aus. Damit müssen Rückspiegelungszellen ODER-Zellen sein. Wenn man möglichst auf allen Ebenen der Informationsverarbeitung, auch im Übergang zwischen Phonetik und Phonologie, gleichartige neuronale Strukturen annimmt, sollten die Zielzellen der Rückspiegelungsverbindungen die ODER-Zellen sein, die aus solchen systematischen Gründen zur Repräsentation der phonologischen Merkmale gehören (zu den Details vgl. Kochendörfer, 2002, und Teil3 „Phonetik/Phonologie“).

Die Punkte 1-4 führen zur Annahme eines Mechanismus der Rückspiegelung, wie er in Abbildung 2.5.3–8 skizziert ist.

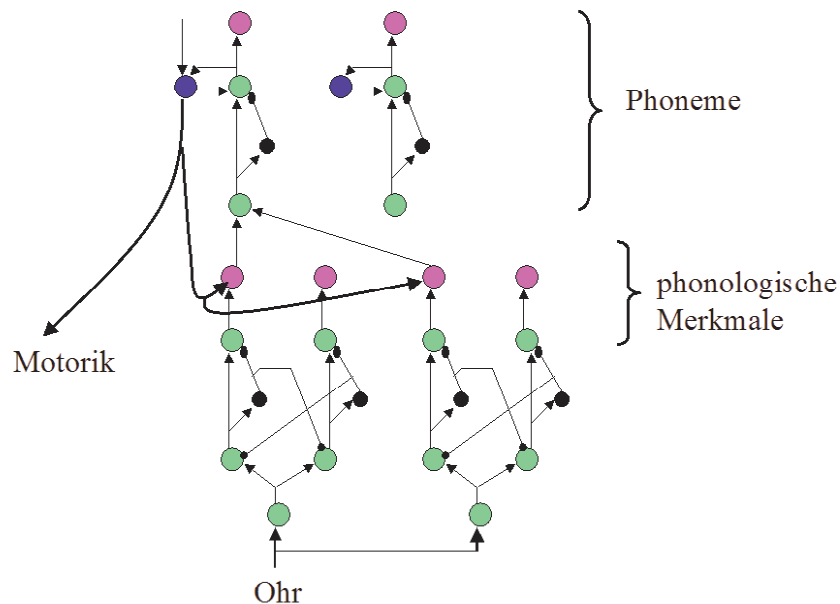


Abbildung 2.5.3–8: Grundzüge einer Rückspiegelungsarchitektur.

Die Rückspiegelungsebene befindet sich bei dieser Architektur an einem Punkt, der so weit „außen“ liegt, wie irgend möglich. In noch größerer Nähe zu den Sinnesorganen gilt Frequenzkodierung, die stabile Lernprozesse ausschließt. Die Rückspiegelungsebene liegt andererseits doch in unmittelbarer Nachbarschaft des Bereichs sprachspezifischer angeborener Verschaltungen.

Dieser Punkt ist wichtig, da besondere, abweichende Strukturen erfordert werden: Um zu gewährleisten, dass nur *eine* Rückspiegelungsebene gebildet wird, darf es potenzielle Verbindungen von P-Zellen auf ODER-Zellen nur an den Rückspiegelungspunkten geben, mit Bezug auf die Sprache also, wie oben festgestellt, bei den phonologischen Merkmalen.

Zusammenfassende Simulation

Für eine zusammenfassende Simulation, die sowohl den Aufbau von Top-down-Strukturen als auch die Verstärkung der Rückmeldungsverbindungen zeigt, soll die Simulationsarchitektur der Abbildung 2.5.3–9 verwendet werden. Sie besteht aus sechs Zellgruppen nach dem Muster der Abbildung 2.5.3–3. Um die hierarchische Anordnung besser darzustellen, sind jeweils zwei Zellgruppen übereinandergestellt, so dass sich drei Spalten ergeben. Die ersten beiden Spalten sind in Szenen-Manier verschaltet, das heißt, dass die Aktivierung der ersten Spalte der der zweiten Spalten zeitlich um einen für Sequenzen typischen Betrag vorausgeht. Es sind Sequenzverbindungen für die Perzeptions- und die Produktionsrichtung vorhanden, so dass der Lernvorgang eine Aktivierung des Produktionstaktsystems als Hilfssystem verwenden kann. Bei der dritten Spalte ist angenommen, dass entsprechende Sequenzverbindungen nicht vorhanden bzw. nicht aktiviert sind, so dass als Lernprozedur nur die Doppelwellenstrategie in Frage kommt.

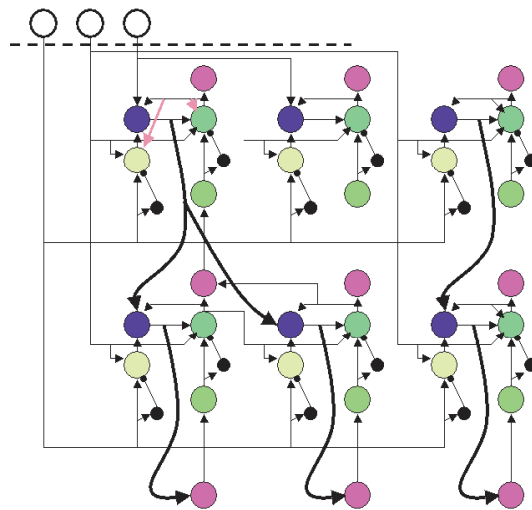


Abbildung 2.5.3–9: Simulationsarchitektur.

Die rot eingetragenen Verbindungen erlauben eine Wiederholung der Aktivierung der betroffenen Einheit, es wird damit ein längeres Andauern der Wahrnehmung und Produktion des entsprechenden Konzepts ermöglicht. Das Problem der Produktion und Perzeption von Längen dieser Art kann hier nicht erschöpfend behandelt werden. Man vergleiche die Teile 3, „Phonetik/Phonologie“, und 5, „Syntax“. Alle Elemente, die nicht mit solchen Verbindungen ausgestattet sind, produzieren nur ein einzelnes Vorkommen.

Da auf der dritten Spalte die Doppelwellenphilosophie dargestellt wird, sind Wiederholungsschleifen bei den sequenzenbildenden Zellen vorausgesetzt, sonst würde die zweite Erregungswelle an diesen Zellen steckenbleiben.

Die Bildschirmdarstellung in der Simulation ist, um die Vergleichbarkeit zu verbessern, in Abbildung 2.5.3–10 wiedergegeben.

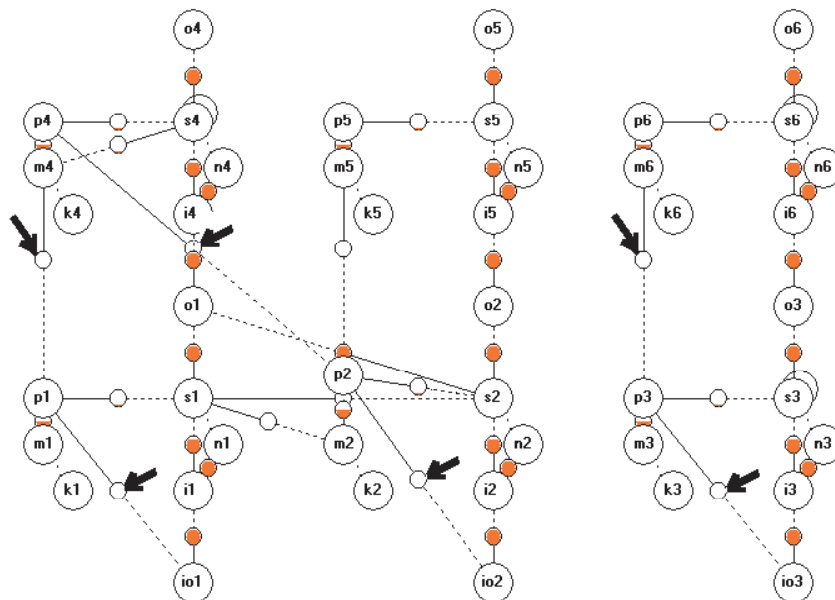


Abbildung 2.5.3–10: Simulation, Bildschirmdarstellung. Die Symbole der top-down orientierten Verbindungsstärken, die sich im Lernprozess verändern, sind mit Pfeilen gekennzeichnet. Die in Abbildung 2.5.3–9 oberhalb der gestrichelten Linie dargestellten Zellen und alle davon ausgehenden Verbindungen erscheinen nicht auf dem Bildschirm, sind aber vorhanden und wirksam.

Simulation:

Lernen von Top-down-Strukturen

Verlauf der Simulation:

Zeittakt	Vorgang
3	Ein sequenzieller Kontext ist durch externen Input gesetzt.
5	Es ist ein Input auf <i>io1</i> erfolgt.
8	Der Produktionstakt ist aktiviert worden.
19	Die sofortige Verstärkung der Rückmeldungsverbindung führt zu einer neuen Erregung von <i>io1</i> . Dieser Effekt entsteht noch zweimal.
65	Es ist ein Input auf <i>io2</i> erfolgt.
140	Die gewünschten Verbindungen sind hergestellt. Die Effekte klingen anschließend allmählich ab.
300 ff.	Testlauf mit der Produktion der Sequenz p1—p2 und den entsprechenden Rückmeldungen.
600 ff.	Verstärkung der Verbindung p6—p3 und der Rückmeldungsverbindung p3—io3 nach dem Doppelwellenprinzip. (Ein Produktionskontext ist nicht gesetzt, daher ist die Hilfssystem-Strategie nicht wirksam.)

2.5.4 Kreative Vorstellungsprozesse

In diesem Abschnitt sollen einige allgemeinere Gesichtspunkte zusammengestellt werden zu einem Thema, dessen ausführliche Ausarbeitung dem Teil 7: „Denken und Formulieren“ vorbehalten ist.

Als „kreative Vorstellungsprozesse“ werden alle Vorstellungsprozesse bezeichnet, die nicht rein reproduktiv sind. Zu den kreativen Vorstellungsprozessen gehört z. B. das oben in Abschnitt 2.5.2 schon einmal erwähnte Probehandeln, und auch das innere Sprechen ist im Allgemeinen nicht Wiederholung von auswendiggelernten Sätzen, also in diesem Sinne kreativ.

Es gibt keine Vorstellungen, die nicht auf Erfahrung beruhen würden. Kreativer Umgang mit Vorstellungen kann also, wie schon oben in Abschnitt 2.5.1 angedeutet, nur heißen, dass vorhandene Elemente neu kombiniert werden, wobei Sinnesprimitive (identisch mit den elementaren neuronalen Strukturen der Sinnesorgane) als Elemente ausgeschlossen sind, das heißt, dass die benutzten Elemente auf einer Ebene angesiedelt sind, auf der schon Lernprozesse möglich sind. Das ergibt sich auch aus der Diskussion des vorigen Abschnitts über die Rückspiegelungspunkte.

Die Funktion von Vorstellungen besteht wahrscheinlich tatsächlich direkt oder indirekt in der Verhaltenssteuerung. Kreative Vorstellungsprozesse kön-

nen zu Gedächtnisspuren (neuentwickelten Konzepten) führen, die irgendwann Verhaltenssequenzen beeinflussen.

Wenn man kreative Vorstellungsprozesse als ein „Probieren“ sieht, also als Vorgang, der nicht oder mindestens nicht notwendig in einem einzigen Schritt abgeschlossen ist, stellen sich die folgenden Fragen:

- Wann kann ein Vorstellungsprozess als abgeschlossen betrachtet werden?
- Welche Mechanismen sind wirksam, solange der Vorstellungsprozess als nicht abgeschlossen gilt?
- Wie sind abschließende Speicherprozesse zu denken?

Die erste Frage setzt einen Überwachungs- und Bewertungsprozess (Monitoring) voraus, die zweite Frage betrifft Prozesse, die eine Umschaltung des aktivierten Bahnsystems bewirken, die dritte Frage thematisiert Konzeptbildungsprozesse, die durch Vorstellungsprozesse ermöglicht werden.

Monitoring

Die Aufgabe eines Monitoring-Prozesses besteht darin, Anhaltspunkte dafür zu liefern, ob ein laufender Vorgang oder dessen Ergebnis bestimmten Anforderungen genügt. Wenn man sich klar macht, dass ein Ergebnis nur in aktiviertem Zustand bewertet werden kann, kann man die Aufgabe letztlich auf die Überwachung laufender Vorgänge eingrenzen. Die Kriterien, die dabei eine Rolle spielen können, ergeben sich sicherlich teilweise aus den Eigenschaften der aktivierten Konzepte. Der Verlauf und der „Wert“ eines Vorstellungsprozesses kann davon abhängen, ob die beteiligten Konzepte positiv oder negativ markiert sind (das heißt, entsprechende Komponenten haben). Diese Spur soll hier nicht weiter verfolgt werden. Ein anderer, hier interessanterer Aspekt ist die Stimmigkeit des Prozesses, also des Vorstellungsgebildes insgesamt und seine Übereinstimmung mit Erwartungen. Die Frage ist, wie diese Form von Kohärenz (Konsistenz), bzw. die unerwünschte Inkohärenz, festgestellt und zur Vorstellungssteuerung verwendet werden kann.

Wenn man sich an das Beispiel von Kuipers oben in Abschnitt 2.4.6 erinnert, bei dem eine unerwartete Ereignisfortsetzung zu einer Irritation führt, wird klar, dass Inkohärenzen nicht nur bei Produktionsprozessen, und auch nicht in erster Linie bei Produktionsprozessen auftreten, sondern zunächst einmal eine Eigenschaft der Perzeption sind. Da Monitoring selbst ja doch

wohl zunächst ein Wahrnehmungsprozess sein muss (Inkohärenz wird wahrgenommen), liegt es nahe, dort anzusetzen und zu fragen, wie die Kohärenz bzw. Inkohärenz eines *Wahrnehmungsprozesses* festgestellt werden kann. Eine der denkbaren Möglichkeiten kann sofort ausgeschlossen werden: die Überwachung des gesamten Verarbeitungsnetzes durch einen parallel strukturierten Monitor. Es gibt keinerlei Anhaltspunkte dafür, wie dergleichen aussehen könnte, obwohl mit solchen Vorstellungen gelegentlich gearbeitet worden ist (vgl. Schade 1992: 101 ff.).

Die Lösung des Problems ergibt sich, wenn man sich überlegt, wie die Kontrolle eines Verarbeitungsprozesses durch Vergleich mit Erwartungen (mit dem Ziel, Übereinstimmungen und Abweichungen zu erkennen) zu denken ist und was passiert, wenn Erwartungen nicht erfüllt werden. Erwartungen werden durch sequenziell hervorgerufene Erwartungspotenziale repräsentiert, die es ermöglichen, dass auf einer Zelle, die ein solches Potenzial trägt, aufgrund eines weiteren Inputs ein Aktionspotenzial entsteht; das ist die Funktion der sequenzenbildenden Zellen, die damit, wie oben in Abschnitt 2.4.6 dargestellt, eine Art Kontextfilter bilden. Die Entstehung eines Aktionspotenzials ist die Voraussetzung für den regulären Fortgang des Verarbeitungsvorgangs.

Man kann jetzt formulieren: Ein Prozess ist auf jeden Fall inkohärent, wenn der Kontextfilter alle aktivierten Verarbeitungsbahnen blockiert. Wenn das passiert, verebbt das in den hierarchisierend dargestellten Simulationsmodellen sichtbare „Hochperlen“ der Aktivierungen auf halbem Weg. Daraus ergibt sich, wenn Verarbeitungsbahnen ein Ende haben, die Möglichkeit, durch Beobachtung dieser Enden festzustellen, ob ein Prozess noch kohärent ist: Es müssen, während ein kohärenter Wahrnehmungsprozess im Gang ist, Aktionspotenziale in gewissen Abständen mindestens an einem der Bahnenenden ankommen. (Es ist wahrscheinlich von vornherein aussichtslos, zu fordern, dass eine bestimmte Zahl von Bahnen aktiviert sein muss.) Das wiederum ist leicht durch eine „Watchdog“-Schaltung zu überprüfen. Diese Schaltung muss Signale abgeben, wenn innerhalb einer gewissen Zeitspanne kein Bahnenende aktiviert ist. Ein (unverbindlicher) Vorschlag für die neuronale Realisierung einer Watchdogschaltung ist in verschiedenen zurückliegenden Arbeiten diskutiert worden, eine genauere Analyse bringt Teil 7: „Denken und Formulieren“. Am einfachsten ist die Situation dann, wenn alle Verarbeitungsbahnen auf einer und derselben Zelle konvergieren. Das ist aber sicherlich keine notwendige Bedingung und man muss mit der Möglichkeit mehrerer oder vieler überwachter Prozessspitzen rechnen.

Aus den vorangegangenen Überlegungen zum Monitoring von Perzeptionsvorgängen wird klar, dass das Monitoring eines Produktionsprozesses über

den Vorgang der Rückspiegelung (der ein Perzeptionsprozess ist) erfolgen kann. Eine direkte Überwachung des Produktionsprozesses ist nicht erforderlich. Man erinnere sich an das Monitoring im Produktionsmodell von Levelt (1989), bei dem auch der Perzeptionsprozess zu diesem Zweck verwendet wird.

Der Produktionsprozess ist damit in Ordnung und kohärent, wenn der zentrale Auslöser der Produktion (und damit der gesamte Produktionsvorgang) in der Perception des Rückspiegelungsprozesses parallelisiert wird. Wenn man beachtet, dass eine Divergenz, die auf zeitlich aufeinanderfolgende Teilprozesse führt, für motorische Phänomene charakteristisch ist, und umgekehrt in der Perception konvergente Verarbeitungsstränge vorherrschen, so dass man in Perzeptionsrichtung mit dem Zusammenmünden von Bahnen auf leicht zu überwachenden Zellen rechnen kann, wird die Eleganz dieser Lösung deutlich. Man beachte auch, dass man sich die Frage stellen kann, wohin die Aktivierung in Perzeptionsprozessen eigentlich geleitet wird. Die Antwort könnte sein: auf Kohärenzüberwachungsstrukturen. Wenn man kreative Vorstellungsprozesse als Trial-and-Error-Prozesse sieht, kann jetzt jedenfalls die Komponente „error“ als zentral durch eine Watchdoginstanz aufdeckbare Inkohärenz verstanden werden.

Umschalten des aktivierten Bahnensystems

Hier geht es um die Frage, wie die Komponente „trial“ in Vorstellungsprozessen (als Trial-and-Error-Prozesse verstanden) aussehen kann. Ein Umschalten von Bahnen, das ja in einem System von Großmutterzellen die einzige Möglichkeit ist, die aktuelle, spezifische Aktivierung von Inhalten zu verändern, ist nur möglich als Veränderung des aktuellen Kontexts. Das kann durch die Wirksamkeit eines Sinnesinputs geschehen. Wenn aber in einem Vorstellungsprozess die Produktion definitionsgemäß zwar mit Rückspiegelung, aber nicht zusammen mit einem externen Input über die Sinnesorgane erfolgt, muss in diesem Zusammenhang eine Kontextveränderung durch einen zentral ausgelösten Vorgang erfolgen.

Es ist klar, dass dieser Vorgang von der für die Erkennung des „errors“ zuständigen Watchdogkomponente ausgehen muss. Eine Simulation mit einer sehr einfachen Architektur, die in Abbildung 2.5.4-1 skizziert ist, soll zeigen, wie das wenigstens grundsätzlich geschehen könnte. Der Gesamtvorgang kann in zwei Teilaspekte zerlegt werden: eine Kontexterweiterung einerseits und eine dadurch möglich werdende neue Kontextfestlegung andererseits. Zur Erweiterung des Kontexts führt eine Art Pseudoproduktion, die durch zwei aufeinanderfolgende Impulse auf einem weitgestreuten unspezifischen System („unspezifisches prädiktives System“, ausgehend von einer

Watchdog-Zelle mit der Bezeichnung *dk*t) ermöglicht wird, das P-Zellen mit zusätzlichen Erwartungspegeln versieht. Diese Erwartungspegel ergänzen die für Produktionsprozesse typische Top-down-Erregung der P-Zellen, die ebenfalls von der Watchdogkomponente ausgeht. Reparaturprozesse bei syntaktischen Störungen in der Sprachperzeption benutzen dasselbe System (vgl. Teil 5: „Syntax“), dabei wird aber nur ein einzelner Impuls erzeugt. Andere Lösungen der behandelten Verarbeitungsprobleme sind denkbar.

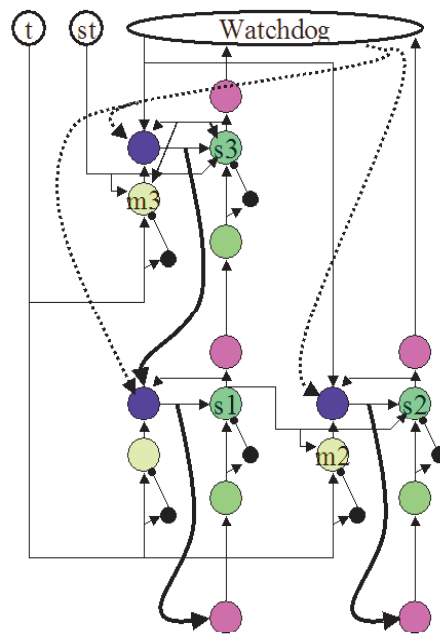


Abbildung 2.5.4–1: Architektur zur Simulation eines Kontextumschaltungsprozesses. Bahnen des „unspezifischen prädiktiven Systems“ sind gestrichelt eingetragen. Die Zellen *t* und *st* und die von ihnen ausgehenden Verbindungen sind auf dem Simulationsbildschirm nicht sichtbar. Von der Watchdoginstanz sind dort nur zwei Ausgänge, die Zellen mit der Bezeichnung *gko* und *dk*t, aufgenommen.

Simulation:
Kontextumschaltung

Verlauf der Simulation:

Zeittakt	Vorgang
1 ff.	Ein relativ enger Kontext wird gesetzt (Erwartungspotenziale auf $m\beta$ und $s\beta$).
10 ff.	Ein Produktionsversuch führt nicht zu einer Rückmeldung.
112 ff.	Die Watchdog-Instanz löst durch eine Art Pseudoproduktion eine Erweiterung des Kontexts aus.
200 ff.	Ein Produktionsprozess aufgrund dieser Erweiterung ist erfolgreich, ein neuer Kontext ist aktiviert.

Der Vorgang benutzt sequenzielle Verbindungen zwischen $s1$ und $s2$ bzw. $m2$, deren Vorhandensein vorausgesetzt wird. Ein neuer Kontext ist selbstverständlich *nicht* ein Kontext, der mit neuen, nicht zuvor durch Lernvorgänge etablierten Konzepten verbunden ist.

Konzeptbildungsprozesse

Konzeptbildung ist eine Leistung instanzenbildender Zellen, die zu dem bottom-up gerichteten Anteil der neuronalen Struktur, also zur Wahrnehmungsseite gehören. Da kreative Vorstellungsprozesse – wie alle Vorstellungsprozesse – zu Ersatzwahrnehmungen führen, ist es selbstverständlich möglich, dass neue Konzepte entstehen können.

Im Einzelnen kann man sich fragen, ob es möglich ist, dass kreative Vorstellungsprozesse eine ausreichend *gleichzeitige* Aktivierung *unterschiedlicher* vorhandener Konzepte auslösen (wie das für Lernprozesse an instanzenbildenden Zellen erforderlich ist), die dann zu einem neuen Konzept für diese Kombination führen würde. Man wird gegenüber der Annahme einer solchen Möglichkeit bei sprachlichen Formulierungsprozessen auf der lautlichen Ebene skeptisch sein, da die angenommene Position der Rückspiegelungsebene eine gleichzeitige Aktivierung unterschiedlicher Phoneme ausschließen sollte. Es ist aber auch auf sprachlicher Ebene möglich, ein vorhandenes Konzept *szenisch* zu erweitern.

Bei anderen Modalitäten, also z. B. visuellen oder taktilen Vorstellungen, ist eine solche Beschränkung durch die Rückspiegelungsbedingungen fraglich. Man beachte in diesem Zusammenhang allerdings, dass auch die visuelle Wahrnehmung eines Bildes ein sequenzieller Vorgang ist. Insgesamt wird man die Möglichkeit der Entstehung neuer Konzepte nicht ausschließen.

Insofern Vorstellungsprozesse zu Ersatzwahrnehmungen führen, können sie ganz selbstverständlich mit lexikalischen Ausdrucksseiten verknüpft werden, genau so wie „echte“ Wahrnehmungen auch. Durch kreative Vorstellungs-

prozesse können damit Wortbedeutungen entstehen, denen keine Sinneswahrnehmungen entsprechen. Es gibt wahrscheinlich keine andere Möglichkeit, Bedeutungen von Abstrakta zu verstehen. Eine andere Frage ist, ob man alle Wörter, deren Bedeutungen durch kreative Vorstellungsprozesse entstehen, als Abstrakta bezeichnen möchte. Angesichts der Schwierigkeiten, die sich aus dem Konzept der Abstraktheit notorisch ergeben, ist es vielleicht sinnvoll, sich per Definition auf eine solche Festlegung zu einigen.

2.6 „Adressierung“ und „Zugriff“ im Kortex

2.6.1 Adressierung

Die Möglichkeit der Adressierung inhaltsneutraler Speicherplätze ist Charakteristikum symbolverarbeitender Architekturen. Aber auch der biologische Kortex stellt einen Raum zur Verfügung, in dem Information untergebracht werden kann. Information ist, gerade in einem lokalistisch organisierten System, nicht eine ungegliederte Einheit, sodass durchaus gesagt werden kann, dass unterschiedliche Informationsbestandteile im Kortex an unterschiedlichen Stellen gespeichert sind bzw. gespeichert werden können. Andererseits ist oben in Abschnitt 2.2.5 festgestellt worden, dass eine Adressierung, realisiert wie in einem symbolverarbeitenden System, zu Strukturen führen würde, die im Kortex nicht zu beobachten sind. Man muss also neu ausloten, was Adressierung im Fall des Kortex bedeuten könnte.

Ein sinnvoller Ausgangspunkt kann sein, dass man sich zunächst überlegt, was sich aus den tatsächlich im Kortex möglichen Strukturen für den Vorgang des „Einspeicherns“ eines neuen Inhalts ergibt. Das Einspeichern muss auch im biologischen Kortex eine Veränderung an einem bestimmten Ort bedeuten.

In gängigen technischen Speichermedien kann eine Information an irgendeiner adressierten Stelle des Speicherraums untergebracht werden. Die Information ändert sich nicht, bzw. jedenfalls nicht notwendig, abhängig von der Stelle, an der das geschieht. Die Information kann in dem Sinne isoliert sein, dass die Nachbarschaft im Speicherraum „leer“ (das heißt: definitionsgemäß leer) ist. Im Unterschied dazu sind im Kortex isolierte Strukturen, die Informationen repräsentieren könnten, nicht möglich. Isoliert heißt: nicht wenigstens indirekt über Bahnen mit Sinnesorganen oder der Motorik bzw. allgemein irgendwelchen Effektoren verbunden.

Manchmal wird gesagt, dass der Kortex ein „inhaltsadressierter“ Speicher sei. Ein inhaltsadressierter Speicher im technischen Bereich kann so aussehen, dass anstelle einer in der Hardware unveränderlich verankerten Adressstruktur Adressen verwendet werden können, die variable Informationen sind, also auch inhaltlichen Charakter haben können. Der Zugriff auf ein Datenelement geschieht mit Hilfe eines solchen inhaltlichen Schlüssels, es muss aber keinen *inhaltlichen* Zusammenhang zwischen dieser Art von Adresse und dem adressierten Datenelement geben. Die Adresse ist nicht Bestandteil des Datenelements, oder muss nicht Bestandteil des Datenelements sein, an dem man interessiert ist. Im Kortex ist auch diese Form der Adressierung nicht möglich.

Es ist ein grundsätzliches Umdenken erforderlich. Man muss sich klar machen, dass jedes Einspeichern von Information im Kortex identisch ist mit einer Bearbeitung, Erweiterung oder Spezifizierung vorhandener Information. Es ist mit angeborenen Inseln von Inhalten (vielleicht identisch mit den primären sensorischen und motorischen Arealen des Kortex) zu rechnen, von denen aus über Lernprozesse Erweiterungen geschehen. Die Erweiterungen müssen in dem Sinne an die angeborenen Inseln anschließen, dass sie neuronal mit ihnen verbunden sind. Wenn man hinzunimmt, dass viele Verbindungen, die ein Neuron ausbildet, relativ kurz bleiben, kann das auch die Konsequenz einer konkret räumlichen Nachbarschaft haben.

„Erweiterung“ heißt also, grob, dass ein neuer Inhalt gebildet wird durch einen vorhandenen Inhalt plus einem Element, das wiederum auf vorhandene Inhalte zurückgeht. Wenn man überhaupt von einem Inhalt als einem Ganzen sprechen möchte, wird er immer auch durch einen Anteil angeborener Verbindungsstrukturen repräsentiert, der mit anderen Inhalten, nicht mit allen, gemeinsam ist.

Diese Verhältnisse haben erhebliche Konsequenzen für die Idee einer Adressierung von Inhalten. Es ist schon allein schwierig, vom „Ergebnis“ eines Adressierungsvorgangs zu sprechen. Es ist irreführend, zu sagen, dass es in der Aktivierung einer Großmuttereinheit oder einer Bahn besteht, denn dadurch würde die Vorstellung begünstigt, dass die Adressierung von dieser Aktivierung getrennt werden kann. Richtig ist allerdings, dass man in diesem Zusammenhang an die Rolle des Kontexts als wesentliches Bestimmungsmoment innerhalb der Adressierung erinnern kann. Ein Kontext besteht in einer Menge von Erwartungspotenzialen, die innerhalb eines bestimmten Zeitfensters innerhalb der Zellen bestimmter Zelltypen vorhanden sind. Der Kontext ist aber, wenn man an den Perzeptionsprozess denkt, in dem Sinne praktisch immer mehrdeutig, als er eine *Menge* möglicher Bahnen charakterisiert, von denen nur eine Auswahl in der Folge wirklich ein Ak-

tionspotenzial führen wird. Im Perzeptionsprozess bestimmen bestenfalls Kontext und Sinnesinput zusammen, welche Wege (Plural!) eine Erregung von außen nehmen wird und wahrscheinlich spielen dabei auch noch Faktoren wie die Länge der Axone und die Geschwindigkeit der Ausbreitung des Aktionspotenzials auf einem spezifischen Axon eine Rolle. Also wird man sagen müssen, dass die Adressierung eines Inhalts letztlich während der Realisierung dieses Inhalts geschieht, und dass Strukturen, von denen man allenfalls sagen könnte, dass sie die Adressierung leisten, identisch sind mit den Strukturen, die die Repräsentation des Inhalts selbst bilden.

Bei Produktionsprozessen, im vorigen Kapitel am Beispiel von Vorstellungen behandelt, ist der Kontext allerdings in einem etwas größeren Ausmaß bestimmend, da es hier ja definitionsgemäß keinen Sinnesinput gibt. Eine gewisse Vergleichbarkeit mit dem Perzeptionsprozess ergibt sich aber, wenn man an den Vorgang der Rückspiegelung denkt und damit an die Voraussetzung der Existenz geeigneter Rückspiegelungsverbindungen, also einer gewissen zusätzlichen Stimmigkeit, die den weiteren Verlauf des Vorgangs mit beeinflusst.

Kontexte als Adressierungsbestandteil werden zunächst automatisch durch vorangehende Prozesse gebildet. Es ist aber möglich, dass die so gebildeten Kontexte nicht einen kohärenten Ablauf ermöglichen; das ist für kreative Vorstellungen in Abschnitt 2.5.4 dargestellt. In diesen Fällen muss man mit einer zentral ausgelösten Kontextumschaltung rechnen, was auch als eine Art von aktivem Adressierungsvorgang gelten kann.

Für die Perzeption liegen die Bedingungen etwas anders als für die Produktion. Es bleibt zwar in der Perzeption der Mechanismus der Kohärenzkontrolle erhalten, die ja auch für den Fall der Produktion über einen Perzeptionsprozess, nämlich den der Rückmeldung, erfolgt. Die Pseudoproduktion, die eine Kontextumschaltung im Fall von Inkohärenz ermöglicht, bleibt ebenfalls in ihrer Funktion gleich. Unterschiede ergeben sich aber durch die Einbeziehung der Sinneswahrnehmungen. Sinneswahrnehmungen sind, anders als Vorstellungsprodukte, nicht vollständig in der Kontrolle des wahrnehmenden Kortex.

Ein Simulationsbeispiel zeigt am besten, mit welchen Vorgängen man rechnen muss. Zu diesem Zweck wird eine Struktur verwendet, die in Abbildung 2.6.1–1 dargestellt ist. Diese Struktur ist um der Übersichtlichkeit willen in bestimmten Punkten vereinfacht: Die ODER-Zellen, die eine feste, überschwellige Verbindung von den sequenzenbildenden Zellen her haben (vgl. oben Abschnitt 2.4.7), sind weggelassen, weil sie nichts zu der darzustellenden Funktion beitragen; die sequenzenbildenden Zellen sind direkt mit instanzenbildenden Zellen verbunden. Die Rückmeldeverbindungen, die über

ODER-Zellen führen müssten, sind ebenfalls direkt mit instanzbildenden Zellen hergestellt. Da ein echter Produktionsvorgang mit einer Aktivität des Produktionsakts nicht vorkommt, sind die M-Zellen und die damit verbundenen hemmenden Zellen weggelassen (zu Rückmeldeverbindungen und Produktionsmechanismen vgl. oben Abschnitt 2.5.3).

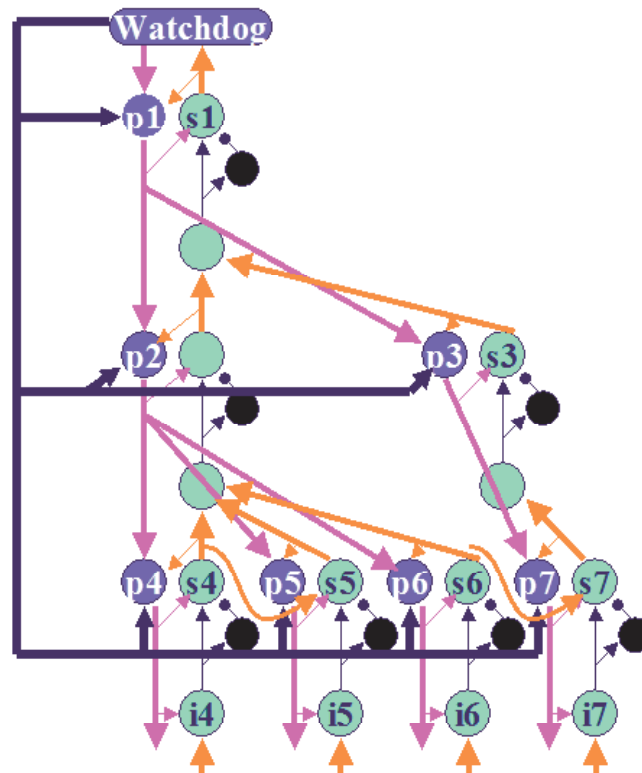


Abbildung 2.6.1-1: Architektur zur Simulation von Kontextumschaltungsprozessen in der Perzeption. Die fett schwarz eingezeichneten Verbindungen bilden das „unspezifische prädiktive System“, sie sind auf dem Simulationsbildschirm nicht sichtbar, aber in der Simulation wirksam.

Simulation:
Kontextumschaltung in Perzeptionsprozessen

Der Simulationsverlauf zeigt einen Wechsel von Kontexterweiterung und Kontextverengung. Die Kontexterweiterung betrifft alle sequenzenbilden-

den Zellen, die im Bereich der von der Zelle *p1* gebildeten Hierarchie liegen. Die einzelnen Prozessschritte sind in der folgenden Verlaufstabelle zusammengestellt:

Zeittakt	Vorgang
4	Ein weiterer Kontext ist gesetzt, der alle sequenzenbildenden Zellen umfasst, außer <i>s5</i> .
30 ff.	Die Zelle <i>i4</i> feuert durch externen Input, der Kontext ist passend, der Watchdog kann ruhiggestellt werden.
110 ff.	Der Input auf <i>i4</i> wird wiederholt, der Kontext wird allmählich auf die mit diesem Input zusammenhängenden Erwartungen beschränkt.
200 ff.	Input auf <i>i6</i> bleibt stecken, weil <i>s6</i> kein Erwartungspotenzial hat.
236 ff.	Die Watchdogzelle <i>gko</i> gibt ein Signal ab, das eine Pseudoproduktion zur Folge hat und zu einer Erweiterung des Kontexts führt.
300 ff.	Erfolgreiche Verarbeitung des Inputs <i>i6</i> , mit entsprechender Verengung des Kontexts bei folgenden Wiederholungen. Das Erwartungspotenzial auf <i>s7</i> wird über die Sequenzverbindung mit <i>s6</i> jeweils aufgefrischt und bleibt dadurch erhalten.
540 ff.	Ein Input auf <i>i7</i> überwindet zwar den Kontextfilter in <i>s7</i> , bleibt aber bei <i>s3</i> stecken.
578 ff.	Die Watchdog-Reaktion löst eine Pseudoproduktion aus, die zu einer erneuten Ausweitung des Kontexts führt.
640 ff.	Der wiederholte Input auf <i>i7</i> ist erfolgreich.

Was man in dieser Simulation beobachtet, ist also nicht ein gezieltes Umschalten auf Kontexte, das dann per Trial-and-Error geschehen müsste, sondern die Auswahl eines mehr oder weniger großen Kontextbereichs. Wenn man die Zeitspanne betrachtet, die für einen *erfolgreichen* Umschaltprozess zur Verfügung steht, ist eine andere Lösung von vornherein unwahrscheinlich.

Das für den Kontextumschaltungsprozess wesentliche „unspezifische prädiktive System“ soll hier noch mit ein paar stichwortartigen Bemerkungen versehen werden, die die besondere Konstruktion rechtfertigen und die Charakterisierung oben in 2.5.4 ergänzen sollen.

- (1) Man kann auf das unspezifische prädiktive System nicht verzichten. Die durch dieses System bewirkten Potenziale können nicht durch eine Erhöhung der Effektivität der top-down gerichteten Verbindungen zwi-

schen P-Zellen ersetzt werden (was vielleicht noch am ehesten denkbar wäre), da eine solche Erhöhung den Produktionsprozess stören würde. Ein normaler Produktionsprozess darf das unspezifische System nicht verwenden.

- (2) Es werden verschieden starke Erregungen gebraucht, die auf P-Zellen einwirken und einen zentralen Ausgangspunkt haben: Erregungen, die nur dann zum Feuern der P-Zellen führen, wenn diese schon ein Erwartungspotenzial haben (wie im Fall syntaktischer Reparaturen bei der Perception gesprochener Sprache) und Erregungen, für die diese Voraussetzung nicht gilt (wie in dem Fall des vorliegenden Kontextumschaltungsproblems). Die erforderliche Pegeldifferenz entspricht dem Unterschied zwischen der Wirkung eines einfachen und der eines doppelten Impulses. Eine Alternative dazu wäre die Verdoppelung des Verbindungssystems mit gleichen Funktionsparametern bei den beiden Teilsystemen.
- (3) Aus Punkt (2) ergibt sich, dass die Watchdog-Einrichtung nicht in jedem Fall einer Inkohärenz einen doppelten Impuls auf dem unspezifischen System auslösen darf. Verarbeitungsprobleme wie die syntaktischen Reparaturen, die weniger zu Bewusstsein kommen, also sozusagen den Normalfall darstellen, kommen, wie oben in Abschnitt 2.5.4 schon vermerkt, mit einem einzigen Impuls aus.

2.6.2 Zugriff

„Zugriff“ auf eine Information meint das Verfügbarmachen dieser Information für einen laufenden Verarbeitungsprozess. In symbolverarbeitenden Modellen bedeutet das nach gängiger Vorstellung die Übertragung von Information in einen zentralen Arbeitsbereich. (Impliziert ist gewöhnlich eine Unterscheidung von passiven und aktiven Strukturen, es besteht die Gefahr der Konstruktion eines Flaschenhalses, der die Möglichkeit zeitparalleler Verarbeitung einschränkt.) Ähnliche Vorgänge werden dann vorausgesetzt, wenn an eine modulare Strukturierung des Kortex gedacht wird und davon gesprochen wird, dass Module Informationen austauschen.

Es ist also, metaphorisch gesprochen, an einen Transport von Information über einen geeigneten Weg von einem Ort an einen anderen Ort gedacht, der natürlich auch Adressierungsvorgänge voraussetzt: Man muss wissen, wo die Information abgeholt und wohin innerhalb des zentralen Arbeitsbereichs sie geliefert werden soll.

Man kann sich nun fragen, an welche Bedingungen ein solcher Transport gebunden ist und ob diese Bedingungen im Kortex gegeben sind. Es ist vor allem zu fragen, welche Eigenschaften das zu transportierende „Objekt“ haben muss. Klar ist, um im Bild zu bleiben, dass es beweglich sein muss. Was hat man sich darunter aber genauer vorzustellen?

Ein einfaches Beispiel, das zur Klärung herangezogen werden kann, ist der Datentransport über eine Telefonleitung. Der Telefonapparat auf der einen Seite der Leitung liefert aufgrund eines akustischen Inputs ein elektrisches Signal, die Leitung wird in einer Zentrale umgeschaltet (Adressierungsvorgang), und das Signal wird am Ende der Leitung von einem Telefonapparat zurückverwandelt in eine akustische Form. Man wird ohne Zweifel von einem Datentransport sprechen und vielleicht auch davon, dass letztlich das Signal das transportierte Objekt ist. Dieses Signal könnte – in einem Gedankenexperiment – ein während einer vorgegebenen Zeitspanne andauernder Spannungspegel fester Höhe sein. Die Information, die sich damit übertragen ließe, wäre buchstäblich minimal. Im realistischen Fall einer tatsächlich existierenden und funktionierenden Telefonleitung ist das Signal wesentlich komplexer, Information ist in einem jeweils spezifischen Signalverlauf kodiert, und die Kodierung ist es letztlich, die es ermöglicht, überhaupt von Datentransport zu sprechen. Es ist selbstverständlich, dass die über eine Telefonleitung übermittelbare Information variabel ist.

Man kann sich die Kodierung auch anders vorstellen, als sie auf einer Telefonleitung realisiert ist. Man muss sich nicht darauf beschränken, nur eine einzige Leitung zu verwenden, sondern man kann mehrere einsetzen, und man kann Muster von gleichzeitig anstehenden Spannungspegeln auf dem Leitungsbündel als kodierte Information verwenden. Das entspricht, wenn man hinzunimmt, dass nur bestimmte Spannungspegel gültig sein sollen, der Technik in handelsüblichen Computern. Wenn man an verteilte konnektionistische Systeme denkt, dann kann man dort einen Informationstransport realisieren, indem man die Output-Neuronen eines Teilnetzes mit den Input-Neuronen eines anderen Teilnetzes so verbindet, dass das Erregungsmuster der Output-Neuronen perfekt kopiert wird (diese Idee wird bei rekurrenten Netzen auch verwendet, siehe Abschnitt 2.2.3 und Abbildung 2.2.3–2). Es macht also kein Problem, bei verteilten Repräsentationen von einem Informationstransport zu sprechen, zusätzlich ist in dem skizzierten Fall die Adressierung unproblematisch. Auch hier ist, wie bei der Telefonleitung, die auf einem und demselben Leitungsbündel zu übermittelnde Information variabel.

Anders ist es bei lokalistischen Systemen. Wir sind, nach der Argumentation in Kapitel 2.3, davon ausgegangen, dass die Repräsentation von In-

formation im Kortex durch die Großmutterzellentheorie annähernd korrekt beschrieben wird. Man beachte nun, dass die Großmutterzellentheorie nicht behauptet, dass einzelne Zellen oder elementare Verschaltungen *in Isolation* eine Bedeutung tragen. Die Bedeutung einer Großmuttereinheit wird konstituiert durch ein Bahnensystem, das sie mit der Sinnesperipherie oder der Motorik verbindet. Wenn von Information die Rede ist, dann muss ein entsprechender Teil des Bahnensystems als Träger gesehen werden. Dazu kommt nun, dass man eigentlich besser nicht sagen sollte, dass die Information von einem Bahnensystem *getragen* wird, sondern vielleicht eher, dass dieses Bahnensystem (nach entsprechenden Lernprozessen, die zu teilweise irreversiblen strukturellen Veränderungen des Systems führen) die Information buchstäblich *verkörpert*.

Es ist offenbar so, dass in lokalistischen Systemen die Vorstellung eines Datentransports auf Schwierigkeiten stößt: Man kann nicht davon ausgehen, dass im Kortex Teile von neuronalen Netzen von einem Ort an einen anderen Ort verschoben werden. Die oben angesprochene „Beweglichkeit“ von Signalen, die Information kodieren, ist nicht gegeben. Es mag sein, dass man vermeiden sollte, überhaupt von Kodierung zu sprechen. Was in einem lokalistischen System möglich ist, ist Instanzenbildung, das kann im einfachsten Fall heißen: Verbindung einer „undefinierten“ Zelle mit einer anderen Zelle, die bereits eine Funktion hat. Dadurch kann eine Art Kopie einer Information hergestellt werden. Es kann auch vorübergehende Kopien (das bedeutet: kurzfristig aufgebaute und wieder abgebaute Verbindungen) geben. Die Frage ist, ob man damit Funktionen realisieren kann, die sonst für einen Informationstransport typisch sind.

Eine typische Funktion eines Datentransports bei mentaler Verarbeitung ist, wie schon angesprochen, das Übernehmen einer Information in eine zentrale Verarbeitungsstruktur (als Arbeitsgedächtnis bezeichnet) und die Rückspeicherung eines Verarbeitungsergebnisses. In einer lokalistischen Architektur könnte man sich vorstellen, dass eine solche zentrale Verarbeitungseinheit über einen Schalter, der Verbindungen verändern kann, mit jeweils gewünschten lokalistischen Konzeptrepräsentationen verbunden ist, wie in Abbildung 2.6.2–1 skizziert. Der Schalter wäre damit verantwortlich dafür, dass Zellen innerhalb der zentralen Verarbeitungsstruktur durch ihre Verbindungen mit definierten Zellen außerhalb dieser Verarbeitungsstruktur eine Bedeutung bekommen, die diesen Zellen entspricht. Damit wird der Datentransport sozusagen ersetzt durch Herstellen einer definierenden Verbindung. Der Schalter müsste über eine Steuerung verfügen, die entscheidet oder Informationen darüber hat, welche Verbindung nun eigentlich hergestellt werden soll, oder aber man denkt – etwas mehr dezentral – daran, dass

eine der vor dem Schalter liegenden Zellen feuert oder zum Feuern gebracht wird, während auch die (bzw. eine) Zielzelle der zentralen Verarbeitungsstruktur mindestens ein erregendes Potenzial hat (beides Funktionen, die dann außerhalb des Schalters liegen).

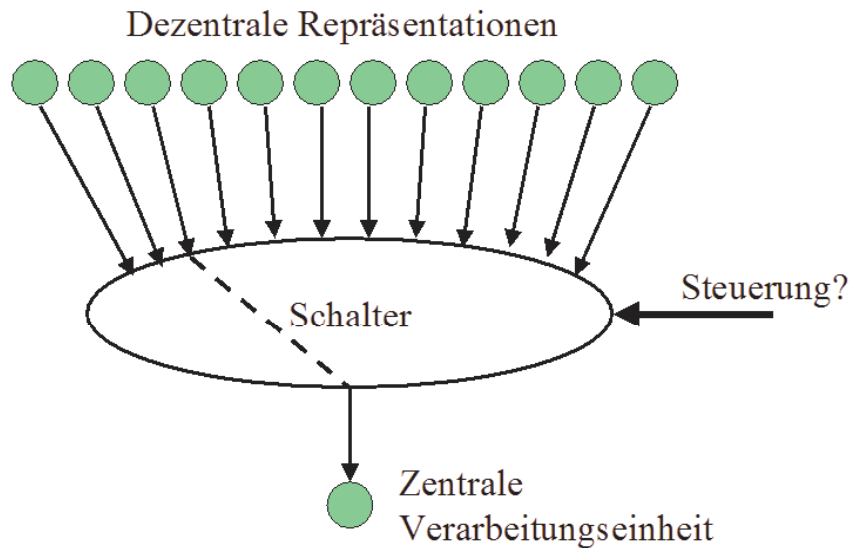


Abbildung 2.6.2-1: Skizze zur Möglichkeit eines Kopiervorgangs für Informationen auf eine zentrale Verarbeitungseinheit unter der Annahme lokalisierender Architekturprinzipien. Die Idee ist, dass Elemente der zentralen Verarbeitungseinheit durch Verbindung mit definierten Zellen außerhalb dieser Einheit vorübergehend eine Bedeutung erhalten. Die Informationen werden über die geschalteten Verbindungen nicht in kodierter Form übertragen.

Ein erstes Problem dieser Idee ist sofort einsichtig: Es können nicht Arbeitsergebnisse auf eine analoge Weise aus dem zentralen Arbeitsbereich heraus übermittelt werden. Arbeitsergebnisse sollten doch wohl darin bestehen, dass Zellen, die eine andere als die ursprünglich eingegebene Information repräsentieren, aktiviert werden. Wie sollen Zellen, die außerhalb des Arbeitsbereichs liegen, diese Bedeutung übernehmen und behalten, wenn der Arbeitsbereich durch den folgenden Teilprozess neu besetzt wird?

Es ist aber auch klar, dass die vorausgesetzte Steuerung des Schalters oder die besondere Aktivationskonstellation der beteiligten Zellen auf erhebliche Probleme stößt. Woher sollen die erforderlichen Informationen kommen?

Wie sollen sie aussehen (unter der Annahme, dass Zellen im Kortex keine Adresseninformation tragen)?

Ein Zugriff auf Daten, der eine Übertragung in einen zentralen Arbeitsspeicher bedeuten würde, ist ganz offenbar im Kortex nicht möglich.

Eine oberflächlich gesehen alternative Form von Zugriff, die in der Literatur diskutiert wird, besteht darin, dass Daten verfügbar gemacht werden durch Aktivierung von ausgewählten Knoten aus einer vorhandenen Menge. Eine Ausprägung dieser Idee ist „spreading activation“: Ein aktivierter Knoten gibt Aktivierung an einen mit ihm verbundenen Knoten weiter. In der Version von Collins & Loftus (1975), die eine weite Verbreitung gefunden hat, kann die weitergegebene Aktivierung unterschiedliche Intensität haben. Von einem Erregungszentrum aus betrachtet nimmt die Aktivierung bei größer werdender Entfernung von diesem Zentrum ab. Mehr oder weniger große Entfernung bedeutet z. B., dass die Erregungsausbreitung über mehr oder weniger dazwischengeschaltete Knoten führt.

Es ist typisch, dass oft unklar bleibt, was durch die Aktivierung eines Knotens eigentlich im Detail bewirkt werden soll. Das gilt z. B. für Versuche, entsprechende Anordnungen als Möglichkeiten für die Modellierung eines Arbeitsgedächtnisses zu sehen. Mit dieser Problematik verwandt ist, dass den einzelnen Knoten Bedeutungen zugeschrieben werden, ohne dass gezeigt wird, wie es dazu kommt, dass ein bestimmter Knoten gerade eine bestimmte Bedeutung hat. Es genügt nicht, dass, wie in entsprechenden grafischen Darstellungen üblich, in die Kreise, die die Knoten darstellen sollen, passende Bezeichnungen geschrieben werden.

Die Definition der Knoten kann (in Perzeptionsrichtung) nur dadurch geschehen, dass entsprechende Verbindungen mit der Sinnesperipherie hinzugegacht werden. Nach dieser „Reparatur“ entpuppen sich die flächenhaft aktivierenden Verbindungen der Netze als Kontextverbindungen, denn die Übertragung von Erregungsintensitäten ist über neuronale Verbindungen in Lernbereichen des Kortex nicht möglich, wie in Kapitel 2.3 gezeigt worden ist. Kontextverbindungen dürfen aber allein und mit einzelnen Impulsen nicht zu einer Aktivierung von Zellen führen, wenn man die Überschwemmung des gesamten Netzes vermeiden möchte, sondern werden nur zusammen mit Perzeptionsprozessen wirksam. Eine sich ausbreitende Aktivierung, die als ein intern gesteuerter Zugriff auf Daten ohne die Einbindung von definierenden Perzeptionsprozessen verstanden werden könnte, ist nicht denkbar.

Wenn man „Zugriff“ generell als Aktivierung sieht und ausschließt, dass es einen Zugriff auf Daten gibt, der einen Datentransport voraussetzen würde,

gilt für den Kortex: Zugriff meint Perzeption und/oder Produktion. Tertium non datur.

2.6.3 Konsequenzen

Wenn es keinen Datentransport im Gehirn gibt, kann auch nicht vom Einspeichern oder Encodieren einer Information an einer bestimmten Stelle des Gehirns gesprochen werden. Es gibt keine „Routen“, auf denen Information weitergegeben wird und keinen Träger der Information. Der Begriff des Signals wird fragwürdig:

„**signal: 1.** Detectable transmitted energy that can be used to carry information. **2.** A time-dependent variation of a characteristic of a physical phenomenon, used to convey information. **3.** As applied to electronics, any transmitted electrical impulse.[...] **4.** Operationally, a type of message, the text of which consists of one or more letters, words, characters, signal flags, visual displays, or special sounds, with prearranged meaning and which is conveyed or transmitted by visual, acoustical, or electrical means.“ (American National Standard, 2001)

Wenn man ein Signal als übertragenen Impuls sieht, ist es identisch mit dem Aktionspotenzial auf einem Axon. Dafür ist letztlich die Definition in Punkt 3 des Zitats zu schwach, weil man doch an eine spezifische Funktion denken muss. Die anderen Definitionen treffen nicht zu, weil dort auf die Eigenschaft verwiesen wird, Information zu tragen. Stichwörter wie „information“ selbst und „data“ in derselben Quelle ergeben ähnlich unbefriedigende Ergebnisse. Es ist offenkundig, dass die gängigen Metaphern zur Beschreibung datenverarbeitender Prozesse beim Versuch einer Anwendung auf das Gehirn versagen.

In dieser Situation ist verständlich, dass die Beobachtungen durch bildgebende Verfahren wie fMRI (siehe Teil 1, Abschnitt 1.1.2) gerne so beschrieben werden, dass eine bestimmte Region des Gehirns an einem bestimmten Vorgang „beteiligt“ sei. Das heißt, es wird diejenige Formulierung gewählt, die am wenigsten über das physikalisch Beobachtbare hinausgeht. Damit werden Aussagen über eine mögliche Zusammenarbeit der Bereiche unterdrückt. Die Rede von der „Beteiligung“ impliziert zwar, wenn man es genau nimmt, häufig die Annahme von Prozessen des Informationsaustauschs zwischen einzelnen Modulen, kann aber auch, wenn man das möchte, neutral gesehen werden. Unproblematisch, was den Informationsaustausch angeht,

ist die These, dass ein Bereich durch einen besonderen Grad der Aktivierung herausgehoben wird, problematisch wird es, wenn von Beobachtungen ausgehend doch wieder Bereiche im Kortex ausgewiesen werden, die mit „CENTRAL EXECUTIVE“, „FOCUS OF ATTENTION“ usw. bezeichnet werden (man vgl. dazu Chein, Ravizza & Fiez, 2003).

Wenn es schwierig wird, die Übertragung von Daten in einen – vielleicht auch nur für eine bestimmte Aufgabenstellung zuständigen – zentralen Arbeitsbereich und ein von dort ausgehendes Abspeichern von Ergebnissen anzunehmen, hat das weitreichende Konsequenzen für die Sprachtheorie. Man kann eine direkte Verbindung herstellen zwischen der Annahme einer zentralen Verarbeitungsinstanz einerseits und der Idee der größtmöglichen Regelhaftigkeit und Redundanzfreiheit der Kompetenz andererseits: Zentrale Regeln und Prinzipien sollen auf möglichst viele Daten anwendbar sein, sowohl die Datenseite als auch die Verarbeitungsseite sind entsprechend strukturiert zu denken. Die methodische Konsequenz für die Arbeitsweise des Linguisten, die im Rahmen der generativen Sprachtheorie eine in allen Entwicklungsphasen gültige grundlegende Rolle gespielt hat, ist die Suche nach immer neuen, noch leistungsfähigeren Generalisierungen.

Wenn Daten aber nicht auf einen Verarbeitungs-Punkt zusammengeführt werden, sondern entsprechend verstreut und an ihren jeweiligen Positionen ortsfest gedacht werden müssen, ist es auch nicht erforderlich und wahrscheinlich nicht einmal möglich, dass Regeln oder Konstruktionsprinzipien nur jeweils in einem Exemplar vorhanden sind. Wenn es aber mehrere bzw. viele Exemplare gibt oder geben muss, ist es auch wahrscheinlicher, dass es zu Abweichungen kommt, und eine Generalisierung – beinahe um jeden Preis – an den Realitäten vorbeigeht. Dass eine gewisse Redundanz der Repräsentationen (nicht nur der von Lernprozessen vorausgesetzten Ausgangsstruktur!) unvermeidlich ist, ist schon oben in Kapitel 2.3 festgestellt worden.

Es gilt in der Tat, was Chomsky in einer 1982 publizierten Diskussion (die eher ein Interview ist) so formuliert hat:

„... it might be a fundamental error to search for too much elegance in the theory of language, because maybe those parts of the brain developed the way they did in part accidentally. For example, what has been so far a very productive leading idea, trying to eliminate redundancy, could be argued to be the wrong move, because in fact we know that biological systems tend to be highly redundant for good reasons. Suppose it does turn out that biological systems are messy, either because of historical accident or maybe because they

work better when they're messy. They may provide many different ways of doing the same thing. If something fails, something else will work. To the extent that that is true. The[sic!] theory of these systems is going to be messy too. If that would be the case, it might be a really fundamental error to be guided too much by an effort to eliminate redundancy in developing explanatory theories.“ (Chomsky, Huybregts & Riemsdijk, 1982: 30f.)

Chomsky war sich also der Problematik bewusst, die im Bestreben besteht, linguistische Theorien redundanzfrei zu halten. Das bedeutet natürlich nicht, dass er diese Leitidee der generativen Sprachtheorie ernsthaft relativieren wollte.

Eine andere Frage ist, warum wir überhaupt dazu neigen, nach „eleganten“ Lösungen zu suchen.