Günter Kochendörfer

Kortikale Linguistik

Teil 4: Lexikon, Morphologie

http://www.cortical-linguistics.de
10. 6. 2005

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 4

Lexikon, Morphologie

Das mentale Lexikon ist eine Komponente, die, wenn man an die prozesshafte Funktion denkt, eine Position zwischen Phonetik/Phonologie und Satzverarbeitung (Syntax wäre als Begriff hier zu eng) einnimmt. Eine gewisse phonetisch/phonologische Vorverarbeitung wird vor Anwendung lexikalischer Informationen im Sprachperzeptionsprozess vorausgesetzt und die Satzverarbeitung setzt ihrerseits Leistungen der Lexikonkomponente voraus.

Wenn man im Hinblick auf die Besonderheiten der mentalen Verarbeitungsprozesse das Lexikon nicht als abgeschlossenes Modul sieht, ergeben sich vielfache, auch auf verschiedene Zeitpunkte ausgedehnte Abhängigkeiten, und lexikalische Prozesse stehen nicht blockhaft vorangehenden und folgenden Prozessen gegenüber. Allerdings läßt sich zeigen, dass die Strukturen, die als lexikalisch zu gelten haben, von den Strukturen in der Phonetik/Phonologie und in der Syntax charakteristisch abweichen, obwohl das nicht bedeutet, dass die allgemeinen, in Teil 2, „Grundlagen", ausgearbeiteten Prinzipien, die sich aus der „Hardware"-Struktur des Gehirns ergeben, außer Kraft gesetzt würden.

Lexikonstrukturen sind in vielfachen Zusammenhängen diskutiert worden. Obwohl dabei oft Bedingungen eine Rolle gespielt haben, die oberflächlich gesehen nicht als mental bezeichnet werden können, kann man solche Diskussionen doch verwenden, um Einsicht in grundsätzliche Probleme zu bekommen. Es ergibt sich eine Liste von Anforderungen, die als Ausgangspunkt für weitere Analysen dienen kann.

Natürlich gelten linguistische Grundsatzunterscheidungen: Es muss behandelt werden, was unter lexikalischen Ausdrucksseiten und lexikalischen Inhaltsseiten zu verstehen ist und wie produktionsseitige und perzeptionsseitige Prozesse und Strukturen miteinander verknüpft zu denken sind. Es muss auch geklärt werden, wie Wörter, Morpheme, Silben und das Konzept des Lexikoneintrags zusammenhängen und was man sich unter Wortarten und syntaktischen Informationen in Lexikoneinträgen vorzustellen hat.

Schließlich ist es wichtig, darzustellen, wie die Zusammenarbeit mit anderen Komponenten der Sprachfähigkeit aussieht. Eine Lexikonstruktur und lexikalische Prozesse sind nur akzeptable Konstruktionen, wenn sie sich problemlos in einen größeren Zusammenhang einbetten lassen.

Inhalt

4.1 Modelle, Probleme
4.1.1 Computerlinguistik und künstliche Intelligenz
4.1.2 Psychologie und Neurolinguistik
4.1.3 Konnektionismus
4.1.4 Zusammenfassende Bewertung
4.2 Anforderungsprofil
4.2.1 Allgemeine Gesichtspunkte
4.2.2 Prälexikalische Repräsentationen
4.2.3 Anforderungen an lexikalische Strukturen (Ausdrucksseiten)
4.2.4 Zusammenhang von Perzeptions- und Produktionsstrukturen
4.2.5 Repräsentation und Wirksamkeit „inhaltlicher“ Strukturen
4.2.6 Effekte in Experimenten
4.3 Lexikalische Ausdrucksseiten
4.3.1 Grundprinzipien der Repräsentation von Phonemsequenzen
4.3.2 Verbände von Ausdrucksseiten
4.3.3 Lernprozesse
4.4 Lexikalische Inhaltsseiten und Modellierung von Produktionsprozessen
4.4.1 Der Weg vom Ausdruck zum Inhalt
4.4.2 Lernprozesse
4.4.3 Mehrdeutigkeit
4.4.4 Produktionsstrukturen: Grundlagen
4.4.5 Produktionsstrukturen: Lernprozesse

4.5 Silben und Morpheme
4.5.1 Silben in lexikalischen Ausdrucksseiten
4.5.2 Lernen von Morphemen: Segmentierung
4.5.3 Das morphologische Lexikon
4.6 Schnittstellen und Konsequenzen
4.6.1 Schnittstellen
4.6.2 Priming-Effekte
4.6.3 Wortfrequenz-Effekt
4.6.4 Kohorten-Theorie
4.6.5 Inkrementelle Produktion und Perzeption

4.1 Modelle, Probleme

4.1.1 Computerlinguistik und künstliche Intelligenz

Lexikonmodelle, die der Computerlinguistik oder der künstlichen Intelligenz zuzuordnen sind, können, wie schon im Abstract dieses Teils 4 festgestellt worden ist, dazu dienen, den Blick auf grundlegende Probleme zu schärfen. Sie unterliegen allerdings spezifischen Hardware-Bedingungen, man kann nicht davon ausgehen, dass es den maschinenunabhängigen Algorithmus gibt, der einmal auf einem handelsüblichen Computer, ein andermal in einem Säugetiergehirn implementiert sein könnte (vgl. dazu insgesamt Teil 1, „Wissenschaftstheoretische Voraussetzungen“).

In der Sicht der Informatik muss das sprachliche Lexikon als ein relativ großer Datenbestand angesehen werden, der ein Suchproblem verursacht, wenn ein Vorgang des Sprachverstehens in einem Computer realisiert werden soll: Das Lexikon muss mit einem Element des sprachlichen Inputs (einem Wort) abgeglichen werden, das heißt, es muss ein Element des Datenbestands identifiziert werden, das zu dem Inputelement passt. Man kann sich im einfachsten Fall einen Datenbestand in Form einer Liste vorstellen, in der lexikalische Ausdrucksseiten entsprechenden Inhaltsseiten (was immer das im einzelnen Fall heißen mag) zugeordnet sind. Wenn die Liste nicht in besonderer Weise strukturiert ist, kann man sich die Suche nach dem richtigen Eintrag so vorstellen, dass einfach vom Anfang der Liste beginnend Element für Element mit dem Inputelement verglichen wird, bis der entsprechende Eintrag gefunden ist. Ein solcher serieller Suchprozess ist relativ aufwendig (das heißt: langsam) und es sind zahlreiche Strategien entwickelt worden, um die Suche effektiver zu machen.

Eine besonders effektive Klasse von Suchalgorithmen nutzt die Tatsache, dass jedes Zeichen in einem handelsüblichen Computer binär kodiert ist, also auch als Binärzahl interpretiert werden kann. Da die Zellen des Speicherraums, der den zu durchsuchenden Datenbestand enthält, binär kodierte

Adressen haben, ist eine naheliegende Idee für den perfektesten, schnellsten, als „direkt“ bezeichneten Zugriff auf das zu dem Inputelement passende Element des Datenbestands die Verwendung der Binärzahl, die dem Eingabeelement entspricht, als Adresse für die gesuchte Information („hashing“). Die Probleme bei dieser Idee sind, dass der Speicherraum in der Regel kleiner ist, als für dieses Verfahren erforderlich (die Binärzahlen können sehr groß werden), und dass zu erwarten ist, dass er nicht effektiv genutzt wird, das heißt, dass er zu einem großen Teil „leer“ sein wird. Also müssen Maßnahmen ergriffen werden, die die Adressen modifizieren. Ein naheliegendes Verfahren ist es, den binären Wert durch die Zahl der verfügbaren Speicherplätze zu dividieren und zur Adressierung den sich ergebenden Rest zu verwenden. Das hat natürlich zur Folge, dass ein Speicherplatz dann ggf. von mehreren verschiedenen Inputelementen adressiert wird. Man muss also vorsehen, dass für diesen Fall eine Teilliste von Lexikonelementen eingerichtet werden kann, die von dem direkt adressierten Speicherplatz aus erreichbar ist, und die dann z. B. seriell zu durchsuchen ist.

Es ist von vornherein klar, dass solche, auf mathematischen Operationen beruhende Verfahren im Kortex nicht möglich sind. Andere Lösungen verwenden spezifische Datenstrukturen in Baumform, um das Suchverfahren zu unterstützen. Eine Klasse dieser Verfahren kommt mit Operationen aus, die in der Feststellung von Gleichheit bzw. Ungleichheit bestehen, ähnlich wie das beim seriellen Suchen in einer Liste auch der Fall ist. Wenn man davon ausgeht, dass ein Inputelement aus einer Kette von Buchstaben besteht, sind Suchverfahren von besonderem Interesse, die vom Kettenanfang ausgehend Buchstabe für Buchstabe auswerten und eine Datenstruktur verwenden, die als „Trie“ bezeichnet wird („digital searching“; vgl. Knuth, 1973: 481 ff., und Ottmann & Widmayer, 2002: 374 f.). Eine Variante davon, die nur mit binären Verzweigungen auskommt, ist in dem Beispiel der Abbildung 4.1.1–1 dargestellt (vgl. Knuth, 1973: 485; Kochendörfer, 1989: 162 f.).

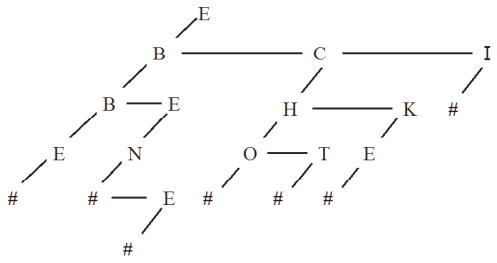

Abbildung 4.1.1–1: Trie für die Wörter *Ebbe, eben, Ebene, Echo, echt, Ecke* und *Ei.*

Der Abgleichvorgang für einen Input ist wie folgt zu denken: Es wird vorausgesetzt, dass es einen Zeiger auf das aktuell zu verarbeitende Zeichen des Inputs gibt. Dieser Zeiger ist zu Beginn auf das erste Zeichen gesetzt. Ein zweiter Zeiger identifiziert den gerade zum Vergleich heranzuziehenden Knoten der Baumstruktur und zeigt zu Beginn auf die Spitze des Baums. Wenn das aktuell zu verarbeitende Zeichen dem aktuellen Knoten der Baumstruktur entspricht, wird der Knotenzeiger auf den Knoten gesetzt, der über den linken Baumzweig erreicht wird, wenn nicht, dann auf den Knoten, der dem rechten Baumzweig entspricht. Außerdem rückt im Erfolgsfall der Zeiger auf den Input um ein Zeichen weiter. Das Verfahren wird fortgesetzt, bis ein # in der Baumstruktur erreicht ist. Mit diesem Knoten ist dann die gesuchte Information verknüpft. Es ist offenbar so, dass Einträge sehr unterschiedlicher Länge für dieses Suchverfahren und die zugrundeliegende Datenstruktur kein prinzipielles Problem darstellen. Eine vom Gesichtspunkt der Sprachverarbeitung aus wertvolle Besonderheit ist, dass beim Scheitern der Suche die größte passende Kette gefunden wird. Dieser Effekt kann dazu verwendet werden, eine Eingabekette in Bestandteile zu segmentieren, die im Lexikon vorhanden sind. Allerdings ist dieses Segmentierungsverfahren nicht deterministisch, das heißt, es können falsche Entscheidungen getroffen werden, die später revidiert werden müssen.

Speicher mit veränderbaren Adressen, bei denen Daten direkt als Adressen verwendet werden können („inhaltsadressierte Speicher“ oder „Assoziativspeicher“) haben, wegen des technischen Aufwands und vielleicht zusätzlich auch wegen der Schwierigkeiten mit variablen Längen sprachlicher Ketten, keine besondere Beachtung bei der Realisierung sprachverarbeitender Prozesse mit Computern gefunden. Man zieht es vor, eine entsprechende Funktionalität durch schnelle Algorithmen (z. B. das oben beschriebene Hash-Verfahren) auf konventionellen Speichern zu erzielen.

Die genannten Strategien setzen voraus, dass sprachliche Ketten in einer Form vorliegen, die es ermöglicht, dass sie tatsächlich direkt mit gespeicherten Mustern verglichen werden können. Davon kann nicht ausgegangen werden, wenn Sprache in akustischer Form, als „fluent speech“, bearbeitet werden soll. Eine ältere Arbeit, die sich mit diesem Problem beschäftigt, ist Klatt (1980). Sie ist auch deshalb für uns interessant, weil es um ein Computersystem geht, das eine Zwischenposition zwischen Künstlicher Intelligenz (im Sinne von ingenieurmäßiger Lösung von Verarbeitungsproblemen) und Psychologie einnimmt. Das entspricht jedenfalls der Einschätzung von Cole (1980: XII):

> „Although the model is presented as an engineering solution to the problem of word recognition, it is intended to be taken seriously as a model of human speech perception.“

Klatt (1980: 246 ff.) sieht acht besondere Problemfelder für die Spracherkennung:

„1. Acoustic-phonetic noninvariance.
2. Segmentation of the signal into phonetic units.
3. Time normalization.
4. Talker normalization.
5. Lexical representations for optimal search.
6. Phonological recoding of words in sentences.
7. Dealing with errors in the initial phonetic representation during lexical matching.
8. Interpretation of prosodic cues to lexical items and sentence structure."
(Klatt, 1980: 247)

Im Einzelnen:

„Acoustic-phonetic noninvariance":

Hier wird von Klatt besonders die Abhängigkeit phonetischer Segmente von ihrer lautlichen Umgebung betont. Auch wenn man kontextunabhängige Merkmale von Lauten extrahieren kann, werden doch offenbar in der natürlichen Sprachwahrnehmung auch kontextabhängige Informationen ausgewertet.

„Segmentation of the signal into phonetic units":

Die phonetische Segmentierung kann nicht im Vorfeld der eigentlichen Identifizierung der phonetischen Einheiten erfolgen, sondern nur in Abhängigkeit von Eigenschaften, die für ein Segment jeweils spezifisch sind. Das führt zu Mehrdeutigkeiten bzw. ist jedenfalls eine nicht zu vernachlässigende Fehlerquelle. „If a way could be found to defer or avoid segmentation decisions, overall system performance might improve."

„Time normalization":

Die Dauer einzelner phonetischer Segmente ist in vielfältiger Weise abhängig von der (variablen) Sprechgeschwindigkeit, von syntaktischen Faktoren und der Betonung. Sie kann nach Klatts Meinung um den Faktor zwei oder drei abweichen. Dazu kommt die phonologisch relevante Unterscheidung von Kürze und Länge. Es ist also erforderlich, relevante von irrelevanten Längenunterschieden zu trennen. Damit steht man aber vor der Situation, dass phonetische Entscheidungen teilweise von lexikalischen und syntaktischen Entscheidungen abhängen, die aber erst getroffen werden können, wenn die phonetischen Entscheidungen bereits vorliegen. Es muss eine Möglichkeit

gefunden werden, Entscheidungen auf einen späteren Zeitpunkt zu verschieben.

„Talker normalization“:

Sprachliche Äußerungen sind in ihrer phonetischen Gestalt sprecherabhängig. Es wirken sich physiologische Differenzen aus, aber auch z. B. die Dialektzugehörigkeit. Zusätzlich wird das Sprachsignal durch äußere Umgebungsbedingungen (Geräusche) oder apparative Umstände (Übermittlung über Telefon) überlagert bzw. verfälscht, ohne dass das allzu dramatische Folgen für das natürliche Sprachverstehen haben muss.

„Lexical representations for optimal search“:

Es ist zu fragen, ob die lexikalischen Repräsentationen so abstrakt sein dürfen, dass sie sowohl für die Sprachproduktion als auch für die Sprachperzeption verwendet werden können. Klatt tendiert von vornherein dazu, mit unterschiedlichen Repräsentationen zu rechnen.

„Phonological recoding of words in sentences“:

Hinter dieser Überschrift, die beeinflusst ist von der generativistischen Vorstellung zugrundeliegender phonologischer Formen, die regelhaft in Oberflächenformen umzukodieren sind, verbirgt sich das für Klatt wichtige Problem der phonetischen Formen bei Wortübergängen (und gelegentlich auch im Wortinnern). An Wortübergängen kann ein Auslaut am Wortende mit dem Anlaut des folgenden Worts zu einem einzigen Segment verschmolzen erscheinen.

„Dealing with errors in the initial phonetic representation during lexical matching“:

Es muss eine Möglichkeit gefunden werden für die Revision von Entscheidungen über phonetische Segmente, die zu nicht korrekten Wortformen führen.

„Interpretation of prosodic cues to lexical items and sentence structure“:

Die Prosodie ist auch schon für den Lexikonabgleich und nicht erst für spätere Prozesse wichtig. Die Entscheidung über die Länge eines phonetischen Segments z. B. ist abhängig von der Betonungskontur.

Klatt übernimmt für sein System LAFS Grundideen aus HARPY, einem System, in dem ganze Sätze, als Lautketten repräsentiert, „verstanden“ werden sollten. Es handelt sich dabei um ein Übergangsnetzwerk, wie es prinzipiell auch sonst aus der Linguistik bekannt ist. Das berühmteste Beispiel ist die „finite state grammar“ in Chomsky (1957), eine andere verwandte Form sind die von Woods (vgl. z. B. Woods 1970) entwickelten Augmented Transition Networks (ATNs). Die Anlehnung an HARPY führt bei Klatt zu

Lexikonstrukturen wie in Abbildung 4.1.1–2 dargestellt. Man erkennt, dass die Wortanfänge so weit wie möglich „zusammengelegt“ sind. Beim Lexikonabgleich wird entlang der Baumkanten jeweils ein entsprechender Pfad von links nach rechts verfolgt. An den Knoten des Baums stehen in Klatts Lexikon Musterspektren („spectral templates“; daher das Akronym LAFS: „lexical access from spectra“), wie in der Abbildung 4.1.1–2 **B** angedeutet. Die Spektren stellen sog. „Diphones“ dar, das heißt Übergänge zwischen zwei phonetischen Segmenten. Man sieht durch den Vergleich von Version **A** mit Version **B**, dass pro phonetischer Einheit zwei Spektren verwendet werden.

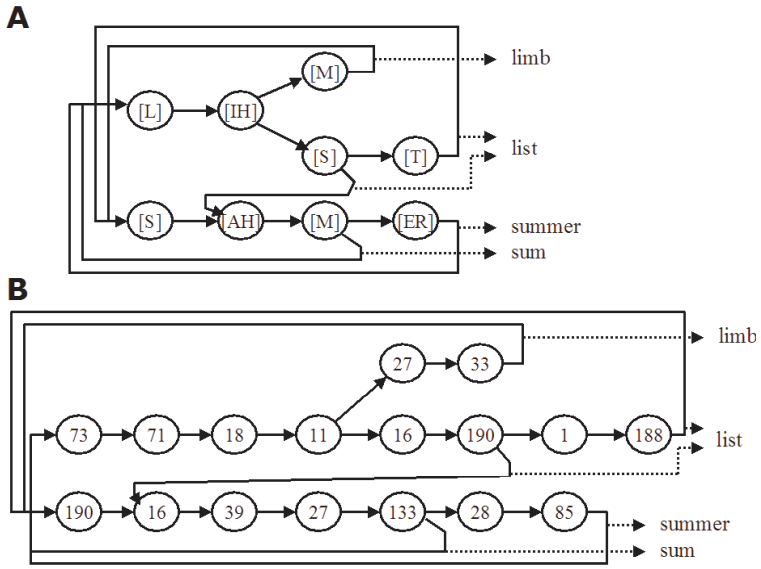

Abbildung 4.1.1–2: Lexikonstruktur in LAFS. **A.** Phonetisches Netz. **B.** Äquivalent mit (durch Nummern bezeichneten) „spectral templates“.

Die gestrichelten Pfeile gehen von den jeweiligen Wortenden aus und deuten den Übergang zu weiterführenden Maßnahmen an. Es gibt genau einen Pfeil für jeden im Lexikon enthaltenen Eintrag.

Eine Besonderheit in LAFS sind die Rückverbindungen von den Wortenden zu den Anfängen der lexikalischen Ketten. Sie dienen dem oben angesprochenen „recoding“ von Segmenten am Übergang von Wortformen in Sätzen. Klatt erklärt das am Beispiel der Wortfolge *list some*, bei deren Aussprache das [t] fehlen kann. Dementsprechend gibt es im Netz eine direkte Verbindung von [S] im Eintrag *list* zu [AH] im Eintrag *some* bzw. *sum* (Klatt,

1980: 262). Dieselben Verbindungen haben noch einen von Klatt ebenfalls gesehenen Nebeneffekt: Sie erleichtern die Segmentierung einer Kette von Phonemen in Wörter. Es muss mit einem neuen Wortanfang nicht nach jeder lautlichen Einheit gerechnet werden, sondern nur, wenn zuvor ein mögliches Wortende (Wortende als Ende einer lexikalischen Kette interpretiert) erreicht ist (vgl. Klatt, 1980: 281). Das Verfahren der Rückverbindungen ist allerdings von Klatt später als möglicherweise zu aufwendig eingeschätzt worden (Klatt, 1989: 196 ff.).

Es sei hier angemerkt, dass die Struktur, mit Ausnahme der Rückverbindungen, praktisch identisch ist mit der oben besprochenen Trie-Struktur. Die Umwandlung des Tries der Abbildung 4.1.1–1 in ein Übergangsnetzwerk ist möglich durch einfaches „Umklemmen“ der waagrechten Verbindungen, wie in Abbildung 4.1.1–3 gezeigt.

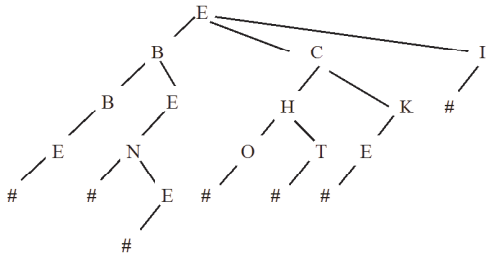

Abbildung 4.1.1–3: Umformulierung des Tries von Abbildung 4.1.1–1 in ein Übergangsnetzwerk.

Wenn man den Graphen um 90 Grad dreht, so dass der Startknoten *E* links liegt, erhält man das für Übergangsnetzwerke typische Aussehen.

In modernen computerlinguistischen Sprachverstehenssystemen werden sog. Hidden-Markov-Modelle zur Repräsentation von lexikalischem Wissen verwendet. Ihre Besonderheit gegenüber einfachen Übergangsnetzwerken ist, dass

- die zeitliche Variabilität des Sprachsignals dadurch modelliert wird, dass der schleifenförmige Übergang zum aktuellen anstelle des Übergangs zum folgenden Zustand im Netzwerk mit einer bestimmten Wahrscheinlichkeit versehen wird und
- die Variabilität eines Outputelements (bzw. eines erwarteten Elements der Eingabedaten) durch Angabe von Wahrscheinlichkeiten für die verschiedenen möglichen Varianten bewältigt wird. (Vgl. Schukat-Talamazzini, 1995: 125 f.)

Die damit verbundenen speziellen Verfahrenstechniken sind computerspezifisch und, wie auch schon die Repräsentation und der Erwerb der erforderlichen Parameter, im Kortex nicht möglich. In unserem Zusammenhang interessant sind einige Überlegungen zu Wortmodellen in lexikalischen Strukturen. Wie bei Klatt werden auch hier Bäume verwendet, die (in der Terminologie von Schukat-Talamazzini, 1995) durch „Verklebungen" der Wortanfänge entstehen. Analoge Vereinfachungen sind aber auch für andere Wortbestandteile, nicht nur die Wortanfänge, möglich. Für die Organisation eines Teillexikons für die Wörter *Ton, tot* und *rot* diskutiert Schukat-Talamazzini (1995: 254) die drei Möglichkeiten der Abbildung 4.1.1–4.

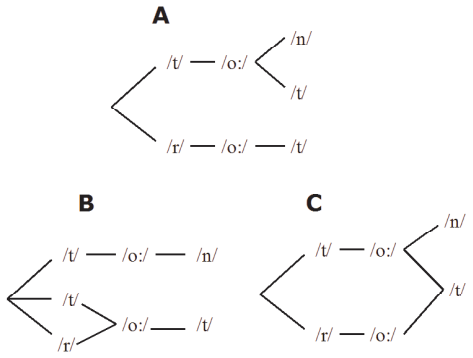

Abbildung 4.1.1–4: Lexikonstrukturen durch unterschiedliche „Verklebungen" von Wortmodellen. **A.** „phonetischer Baum". **B.** „inverser phonetischer Baum". **C.** „phonetischer Rhombus". Nach Schukat-Talamazzini (1995: 254).

Man beachte, dass die gestrichelten Verbindungen an den möglichen Pfadenden, wie sie bei Klatt angedeutet sind (Abbildung 4.1.1–2), bei **B** und **C** von Abbildung 4.1.1–4 keinen Sinn hätten. Die erreichte Einsparung von Lexikonstruktur muss abgewogen werden gegen den zusätzlichen Aufwand, der erforderlich ist, wenn man eindeutig festhalten möchte, welcher Pfad durch das Lexikon bei der Bearbeitung eines Inputs tatsächlich begangen worden ist.

Eine andere interessante Diskussion betrifft die Entscheidung für die Form der Elemente der Wortketten. Ein Problem ist nicht nur die zeitliche, sondern auch die kontextabhängige Variabilität von phonetischen Segmenten. Die Effektivität der Worterkennung kann verbessert werden durch Einführung von kontextabhängigen Phonemen, z. B. von Triphoninventaren. Für die Beispiele *Ton, tot* und *rot* sind erforderlich: *#/t/o:, #/r/o:, t/o:/n,*

t/o:/t, r/o:/t, o:/n/# und *o:/t/#* (gemeint ist, dass das zwischen Schrägstrichen stehende Phonem durch den angegebenen Kontext spezifisch modifiziert ist). Man beachte, dass diese Technik nicht, wie bei einigen konnektionistischen Modellen, zur Kodierung der Sequenz dient!

Anstelle von phonemischen Elementen oder ergänzend dazu können Silben oder ganz allgemein Hierarchien von Zerlegungen verwendet werden. Schukat-Talamazzini (1995: 177) führt als Beispiel eine Zerlegung des Ortsnamens *Dresden* an, die hier in leicht veränderter Form wiedergegeben ist:

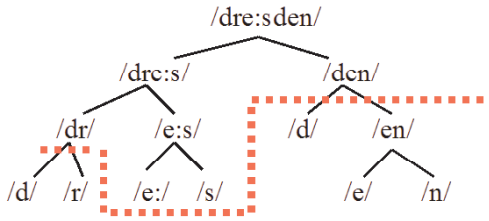

Abbildung 4.1.1–5: Hierarchische Strukturierung von Lexikoneinträgen. Eine mögliche Kette für den Lexikonabgleich ist durch die gestrichelte Linie angedeutet.

Ein Vorteil besteht darin, dass wiederkehrende komplexere Elemente als Einheiten fest definiert werden können und nur ein Rest aus phonemischen Elementen zusammengesetzt werden muss.

4.1.2 Psychologie und Neurolinguistik

Das Interesse bei den in diesem Abschnitt behandelten Modellen liegt weniger bei einer präzisen apparativen Ausformulierung (manchmal als „Implementierung“ bezeichnet), als vielmehr bei der Entwicklung prinzipieller Vorstellungen, ausgehend von Ergebnissen psycholinguistischer Experimente und/oder orientiert an Beobachtungen von Sprachstörungen unterschiedlichster Art.

Für den Verstehensprozess werden in Literaturüberblicken (vgl. Frauenfelder & Floccia, 1999; Massaro, 1994; Lively et al., 1994) in der Regel drei Modelle genannt, die in diese Gruppe gehören: Das „autonome serielle Suchmodell“ von Forster, das „Logogen-Modell“ von Morton und das „Kohortenmodell“ von Marslen-Wilson. Wenn man den Produktionsprozess hinzu-

nimmt, muss zusätzlich mindestens das Produktionsmodell von Levelt (bzw. Roelofs) herangezogen werden.

Forster, autonomes serielles Suchmodell

Der für Forsters „autonomes serielles Suchmodell“ zentrale psycholinguistische Aufhänger ist der sogenannte „Frequenzeffekt“: Bei lexikalischen Entscheidungsaufgaben (und auch bei anderen Designs) kann beobachtet werden, dass häufigere Wörter zu kürzeren Reaktionszeiten führen. Lexikalische Entscheidungsaufgaben bestehen darin, dass für meist visuell dargebotene Stimuli entschieden werden soll, ob sie Wörter einer Sprache darstellen, oder nicht. Dieser Frequenzeffekt wird in Forsters Modell so erklärt, dass das mentale Lexikon Subsets enthält, die in sich nach Häufigkeit geordnet sind und seriell durchsucht werden, daher die Bezeichnung als „serielles Suchmodell“. Das Attribut „autonom“ erklärt sich so, dass keine Interaktion verschiedener Verarbeitungsebenen, also auch nicht eine Rückwirkung höherer Verarbeitungsstufen auf den Lexikonabgleich, vorgesehen ist.

Der gesamte Prozess des Lexikonabgleichs bei visuellen Stimuli sieht, etwas vereinfacht, so aus (nach Bradley & Forster, 1987: 108 f.):

Aus dem visuellen Stimulus werden zunächst orthographische Eigenschaften abgeleitet, die Basis für die Bildung eines Zugriffskodes („access code“) sind. Zur Gewinnung des Zugriffskodes können zum Beispiel Präfixe getilgt werden, oder es wird nur die erste Silbe eines Worts beibehalten.

Der nächste Schritt besteht in der Auswahl eines Lexikon-Subsets (diese Subsets werden als „bins“ bezeichnet).

> „The access code has its first use in the selection of an appropriate subset of the lexicon, over which a detailed search is to be made. Lexical subsets have predetermined boundaries (that is, they are fixed bins), and bin selection is given by a hashing operation on the access code.“ (Bradley & Forster, 1987: 108)

Innerhalb der Subsets wird seriell gesucht, „under an order of encounter which depends on relative frequency, until a match with the access code is found“ (ebd.).

Anschließend wird überprüft, ob die orthographischen Spezifikationen des gefundenen Eintrags mit den vollständigen Eigenschaften des visuellen Stimulus übereinstimmen. Wenn das der Fall ist, kann die mit dem Eintrag verknüpfte syntaktische und semantische Information für die dafür zuständigen Prozesse verfügbar gemacht werden.

Es ist klar, dass dieser Vorschlag von Forster sehr stark an computerlinguistischen Vorbildern orientiert ist und es ist zu fragen, ob der Frequenzeffekt die von Forster behaupteten Strukturen und Prozesse ausreichend begründet. Andererseits mag Forster (1976) Recht haben mit der Feststellung, dass die oben 4.1.1 beschriebenen baumförmigen Lexikonstrukturen („letter-based decision-trees“) Schwierigkeiten mit der Darstellung des Frequenzeffekts haben. Unkritisch ist, dass „Frequenz“ personbezogen gesehen werden muss und vielleicht auch Unterschiedliches meinen kann.

Morton, Logogenmodell

Das Modell von Morton ist mehrfach weiterentwickelt worden, wobei die „Logogen“-Funktion immer mehr in den Hintergrund des Interesses rückt. Es geht bei den Neuerungen hauptsächlich um die Aufspaltung in modalitätenspezifische Teilsysteme (Teillexika: orthographisches Inputlexikon, phonologisches Inputlexikon, orthographisches Outputlexikon und phonologisches Outputlexikon), Spezifizierung von Verarbeitungsbahnen und von Komponenten, die prä- und postlexikalisch wirksam sind.

In der ursprünglichen Fassung (Morton, 1969) besteht das mentale Lexikon aus einer Menge von „Logogenen“, das sind Einheiten, die jeweils für ein Wort stehen, und die für die Entscheidung zuständig sind, ob das betreffende Wort dem Input entspricht. Die Logogene arbeiten zeitparallel, aufgrund der in ihnen enthaltenen Sets von „Attributen“ zur Orthographie, Phonologie und Semantik.

Die Modellierung der Logogen-Funktion ist deutlich an der Neurobiologie orientiert. Logogene können aufgrund der Menge passender Informationen mehr oder weniger stark aktiviert sein. Wenn die Aktivation einen Schwellenwert erreicht, „feuert“ das Logogen, das heißt, die in ihm repräsentierten Informationen werden für die Weiterverarbeitung verfügbar. Wortfrequenzeffekte werden dadurch modelliert, dass Logogene häufigerer Wörter niedrigere Schwellen haben. Die Schwelle eines Logogens wird auch dann erniedrigt, wenn es gefeuert hat. Damit sollen die kürzeren Reaktionszeiten bei Wiederholungen von Wörtern erklärt werden. Über die Interaktion mit einem Kontextsystem (später „cognitive system“) kommen Kontext- und Priming-Phänomene zustande.

Die Vorteile des Logogen-Modells gegenüber Forsters seriellem Suchmodell sind deutlich, trotz der Argumente, die Forster (1976) seinerseits gegen das Logogen-Modell ins Feld führt: Der Frequenzeffekt kann einfacher erklärt werden, auch weitere experimentelle Daten werden besser gespiegelt, das gilt z. B. für Kontexteffekte. Der Gesamtprozess ist nicht mehr so leicht durch Hinweis auf den erforderlichen Zeitbedarf für die serielle Suche anzugreifen.

Die Idee der Aktivationsstärke spezialisierter Einheiten ist in unterschiedlichen Modellen und für unterschiedliche Funktionen wieder verwendet worden.

Marslen-Wilson, Kohortenmodell

Für das Kohortenmodell von Marslen-Wilson sind zunächst zwei Entwicklungsstufen zu unterscheiden, eine ältere, die in den klassischen Texten Marslen-Wilson & Welsh (1978) und Marslen-Wilson & Tyler (1980) dargestellt ist, und eine jüngere, die durch einschneidende Veränderungen gekennzeichnet ist und in Marslen-Wilson (1987) begründet wird. Als dritte Schicht muss das konnektionistische Modell von Gaskell & Marslen-Wilson (1997) angesehen werden, auf das im folgenden Kapitel kurz eingegangen wird.

Ausgangspunkt für Marslen-Wilson und Mitarbeiter sind psycholinguistische Experimente, in denen gezeigt wird (werden sollte), dass die Wortverarbeitung im Verstehensprozess nicht „seriell“ erfolgt (das heißt hier: durch autonome, bottom-up aufeinanderfolgende Verarbeitungskomponenten, von denen der formale Lexikonabgleich eine ist), sondern dass immer mit der Rückwirkung höherer Verarbeitungsergebnisse auf die Entscheidungen auf niedrigerer Ebene, also mit einem komplexen Zusammenwirken der Verarbeitungskomponenten zu rechnen ist. Die Annahme autonomer Komponenten führt unter anderem dazu, dass die Worterkennung abgeschlossen sein muss, *ehe* semantische Einflüsse wirksam werden können. Dass semantische oder allgemeiner: Kontexteinflüsse überhaupt wirksam werden, ist nicht der kritische Punkt, der kritische Punkt ist, *wann* sie wirksam werden.

Versuchspersonen können schon korrekt reagieren, wenn sie zwei Phoneme eines Wortes wahrgenommen haben. Das kann nicht allein durch die lautliche Basis erklärt werden, sondern es müssen zusätzlich Kontexteinflüsse die Entscheidung für eines der gleich-anlautenden Wörter ermöglichen. Auf der anderen Seite kann es auch nicht allgemein so sein, dass immer die ersten beiden Phoneme zu einer Kandidatenauswahl führen, die dann semantisch entschieden wird (das wäre wieder die autonome Aufeinanderfolge von zwei Prozesskomponenten).

> „These arguments bring us to the point where we will have to turn to a quite different type of processing model. That is, a distributed processing model in which recognition is mediated by a large array of individual recognition elements, each of which can integrate sensory and contextual information in order to determine whether the word it represents is present in the signal.“ (Marslen-Wilson & Tyler, 1980: 29)

Aus dieser Grundvorstellung heraus wird der folgende Ablauf konstruiert: Die „pattern-matching capacities“ der parallel arbeitenden Lexikonelemente bewirken, dass, solange nur der Wortanfang wahrgenommen wird, eine Menge von Elementen aktiv wird, die diesem Anfang entsprechen. Diese Menge von Wortkandidaten wird als Anfangskohorte („word-initial cohort“) bezeichnet. In der Folge wird diese Anfangskohorte sukzessive reduziert, bis nur noch ein einziger Wortkandidat übrig geblieben ist.

Die Reduktion der Anfangskohorte kann auf verschiedene Weise erfolgen. Wenn nur der phonologische Input verfügbar ist, gilt:

> „As more of the signal is heard, it will tend to diverge from the internal specifications of more and more of the members of the word-initial cohort. These elements will detect the resulting mismatches, and will therefore drop out of consideration as candidates for recognition. This process will continue until only one candidate is left whose properties still match the signal.“(Marslen-Wilson & Tyler, 1980: 31)

Es ist klar, dass, ergänzend zu diesem rein vom phonologischen Input getriebenen Prozess, im Sinne von Marslen-Wilsons Grundanliegen nun auch noch ein Kontexteinfluss dazu dienen kann, den Umfang der Kandidaten-Kohorte zu reduzieren. Die Lexikonelemente müssen als Spezialisten gesehen werden, die sowohl phonologische als auch inhaltliche Informationen beliebiger anderer Art (Kontextinformationen) verarbeiten können. Das Ergebnis ist ein den jeweiligen Umständen des Verstehens entsprechender optimaler Prozess der Worterkennung, der zu einem möglichst frühen Eintreten des „recognition point“ führt.

Der „recognition point“ ist, anders als das von Mortons Logogen-Modell vorhergesagt würde, nicht nur abhängig von den Eigenschaften vorhandener Wörter, sondern auch davon, welche Wörter nicht vorhanden sind (Marslen-Wilson, 1987: 83).

Die Kritik, die an dem frühen Kohortenmodell von 1978 und 1980 geübt worden ist, hat später zu einigen nicht undramatischen Veränderungen geführt. Der wesentliche Aspekt wird von Marslen-Wilson so formuliert:

> „I will argue, in particular, that the cohort model has to move away from its binary concept of information and decision, where candidates are either in the cohort or out of it, towards a more fluid form of organisation, incorporating the concept of activation.“ (Marslen-Wilson, 1987: 90)

Das hat im Einzelnen die folgenden Konsequenzen für den Vorgang des Lexikonabgleichs als Ganzes:

- Dem Frequenz-Effekt, der die Basis des Modells von Forster war, wird dadurch Rechnung getragen, dass der Aktivations-Level für häufigere Wörter leichter erhöht werden kann als der für seltenere (Marslen-Wilson, 1987: 93).
- Die Bildung einer zu kleinen Anfangskohorte z. B. unter Rauschen und ein zu rigider Abbau durch akustischen Input wird durch Zulassen nicht ganz vollkommener Kandidaten und durch Einführung einer Inputrepräsentation verhindert, die mit subphonemischen Elementen arbeitet.
- Schließlich wird, um allzu drastische Reduktionseffekte zu vermeiden, die Idee eines top-down wirksamen Kontexteinflusses auf der Ebene der „actual lexical recognition units" aufgegeben. Er wird ersetzt durch einen Wettbewerbsprozess der Wort-Kandidaten auf höherer Ebene.

Einige dieser Korrekturen erinnern an Eigenschaften des konnektionistischen Modells TRACE (siehe unten Abschnitt 4.1.3).

Levelt/Roelofs, Sprachproduktionsmodell

Ein großer Teil von Levelts „Speaking" (Levelt, 1989) ist Problemen gewidmet, die mit der Rolle des mentalen Lexikons innerhalb der Sprachproduktion zu tun haben. Es wird dargestellt, wie das mentale Lexikon mit anderen Komponenten des Sprachproduktionsprozesses, der grammatischen und phononologischen Enkodierung interagiert und wie der lexikalische Informationsbestand genutzt wird. Wenn man akzeptiert, dass es die Funktion des Lexikons im Verstehensprozess ist, eine lexikalische Ausdrucksseite mit Inhalten zu verbinden, wird man für die Produktion von Äußerungen annehmen, dass in diesem Fall Inhalte der Ausgangspunkt sind und es darum geht, diesen Inhalten Ausdrücke zuzuordnen. Die Aktivierung der korrekten Ausdrucksseite muss über das inhaltliche (zu äußernde) Konzept erfolgen. Da Wörter nicht nur phonologische, sondern auch syntaktische Eigenschaften haben, ergibt sich eine Zweiteilung des Prozesses. Es wird zuerst das Lemma, das die syntaktische Information enthält, verfügbar gemacht und in einem zweiten Schritt, vom Lemma über einen Zeiger adressiert, die morphologische und phonologische Information.

In Levelt (1989) ist die Adressierungsfunktion des Inhalts grob so gesehen worden, dass der als Komponentenstruktur repräsentierte zu äußernde Inhalt mit den im Lexikoneintrag niedergelegten Inhaltsbeschreibungen der einzelnen Lexikonelemente zeitparallel verglichen werden sollte. Diese Idee

führt zu einem Problem, das als „Hypernymproblem“ diskutiert worden ist. Es besteht darin, dass ein spezifischeres Konzept, das ein allgemeineres Konzept enhält, auch das zu diesem allgemeineren Konzept passende Lemma aktivieren müsste. Also sollte das Konzept LASTWAGEN nicht nur das Lemma *Lastwagen*, sondern auch das Lemma *Auto* aktivieren, unter der Voraussetzung, dass das Konzept AUTO vollständig in dem Konzept LASTWAGEN enthalten ist. Es ist aber offenbar nicht so, dass in einem solchen Fall die Äußerung des spezifischeren Konzepts beim natürlichen Sprechvorgang irgendwelche Schwierigkeiten verursachen würde. Levelts eigene Lösung des Problems auf dem Stand von 1989 ist als Bündel von Prinzipien formuliert und nicht leicht modellhaft auszubuchstabieren (Levelt, 1989: 212 ff). Eine Alternative ist von Bierwisch & Schreuder (1992) vorgeschlagen worden, sie besteht darin, dass angenommen wird, dass hemmende Verbindungen von den spezielleren zu den allgemeineren Konzepten den Konflikt verhindern.

Die schließlich in der Levelt-Schule akzeptierte Lösung besteht aber darin, dass die Vorstellung von der Komponentenstruktur des zu versprachlichenden Konzepts aufgegeben wird (Roelofs, 1992; Roelofs, 1997; Levelt, Roelofs & Meyer 1999; Levelt 2001). Konzepte sind Ganzheiten, die eine bestimmte Aktivation annehmen können. Diese Einheiten adressieren sozusagen direkt die zugehörigen Lemmata, es gibt kein Hypernymproblem und die „lemma selection“ wird zu einem relativ einfachen Vorgang.

Levelt (2001: 13466) beschreibt diesen Vorgang so:

> „Given a target concept, such as LEFT or HORSE, the subject will proceed to select the corresponding lemma. Let us assume that, in a picture-naming experiment, HORSE is the subject's target concept [...]. It spreads part of its activation to the lemma node *horse,* which is the one to be selected. HORSE also sends part of its activation to semantically related concept nodes, such as those for ANIMAL and GOAT. In turn, these spread part of their activation to their lemma nodes, *animal* and *goat*. Hence, the lemma *horse* is selected under competition [...].
>
> After selection of the lemma, its syntactic properties are accessible for further processing.“

Nur ein selektiertes Lemma gibt dann Aktivation an den ihm zugeordneten phonologischen Kode weiter. Dieser Vorgang ist frequenzabhängig: „Accessing the code is slower for low-frequency than for high-frequency words“ (Levelt, 2001: 13467).

Der phonologische Kode wird als Ganzes aktiviert, nicht in Teilen, und er enthält keine Silbenstruktur. Die Silbenstruktur, die über Morphemgrenzen

hinweg gehen kann, wird einem Ausdruck anschließend unter Aktivierung eines sprachspezifischen Silbeninventars aufgeprägt.

Charakteristisch für dieses Modell ist die Spreading-activation-Technik in einer Variante mit spezifizierten Kanten, das heißt, es gibt z. B. Kanten, die Label wie „is a“ oder „has a“ usw. tragen. In dieser besonderen Behandlung der Kanten besteht der wesentlichste Unterschied zu den Modellen, die als „konnektionistisch“ (im engeren Sinne) klassifiziert werden.

4.1.3 Konnektionismus

Von dem Bild ausgehend, das in Teil 2, „Grundlagen“, besonders Abschnitt 2.2.3 von konnektionistischen Modellen entworfen worden ist, müßte man annehmen, dass der Lexikonabgleich, weil es sich ja mindestens annähernd um einen Mustervergleich handelt, ein für die konnektionistische Modellbildung besonders fruchtbares Gebiet ist. Die Details sehen allerdings doch nicht so günstig aus.

Ein einfaches Modell eines konnektionistischen Lexikons ist in der Abbildung 2.2.3–3 des Teils 2 als Beispiel für ein Netz mit lokalistischen Repräsentationen angegeben, diese Abbildung wird hier der Bequemlichkeit halber wiederholt.

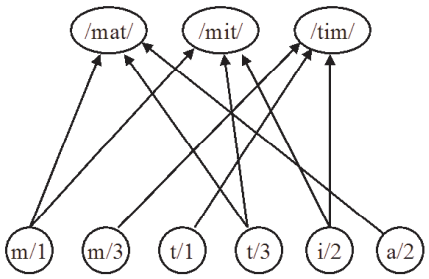

Abbildung 4.1.3–1: Einfaches konnektionistisches Lexikon.

Das Prinzip ist, dass in einer Eingabeschicht pro Laut und pro Position im Wort ein künstliches Neuron vorhanden ist und dass aufgrund von gleichzeitiger Aktivierung entsprechender Neuronen der Eingabeschicht ein Neuron der Ausgabeschicht eine herausgehobene Aktivation hat.

Die Besonderheit besteht in der Behandlung der Reihenfolgeinformation. Das dafür verwendete Verfahren liegt einigen gängigen Lexikonmodellen zugrunde, so auch dem frühen „Interactive Activation“-Modell (im Folgenden

abgekürzt: IA-Modell) von McClelland & Rumelhart (1981). Dieses Modell beschreibt den Leseprozess für Wörter, die aus einem Zeichensatz aufgebaut sind, dessen Buchstaben aus einfachen Strichmustern bestehen, wie man das bei manchen Anzeigen von elektronischen Geräten findet. Die Abbildung 4.1.3–2 erläutert das Prinzip dieser Bustabendarstellung.

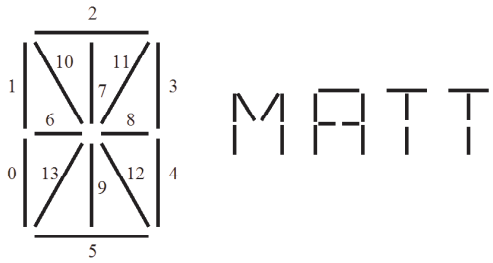

Abbildung 4.1.3–2: Zerlegung eines Eingabebuchstabens (nach McClelland & Rumelhart, 1988: 225). Die Zahlen bezeichnen Positionen in einem Bitmuster, siehe unten.

Die Inputschicht für das konnektionistische Netz wird durch Einheiten gebildet, die einzelne Striche der Buchstabenform repräsentieren. Diese Merkmalseinheiten sind gedoppelt: Es gibt je eine Einheit für das Vorhandensein und eine für das Fehlen eines Strichs. Details werden weiter unten besprochen.

Die zweite Schicht des Netzes wird durch Buchstabeneinheiten gebildet, die dritte durch Worteinheiten. Zwischen den Einheiten sind folgende Verbindungen vorhanden (nach McClelland & Rumelhart, 1988: 210 f.):

- *Verbindungen von Merkmalen zu Buchstaben:* Merkmalseinheiten haben erregende Verbindungen zu allen Buchstaben ihrer Position (!), in denen sie vorkommen und hemmende Verbindungen zu allen Buchstaben ihrer Position (!), in denen sie nicht vorkommen.
- *Verbindungen von Buchstaben zu Wörtern:* Buchstabeneinheiten haben erregende Verbindungen zu allen Wörtern, in denen sie an der entsprechenden Position vorkommen und hemmende Verbindungen zu allen Wörtern, in denen sie an der entsprechenden Position nicht vorkommen.
- *Verbindungen von Wörtern zu Buchstaben:* Worteinheiten haben (rückwirkende) erregende Verbindungen zu allen Buchstabeneinheiten, die zu ihrer Erregung beitragen.

- *Verbindungen innerhalb einer Ebene:* Alle Einheiten innerhalb der Buchstabenebene, soweit sie an einer bestimmten Position konkurrieren, sind hemmend verknüpft. Alle Worteinheiten der Wortebene sind ebenfalls hemmend verknüpft.

Wenn man die Strukturen des IA-Modells auf unser einfacheres Beispiel in Abbildung 4.1.3–1 überträgt, ergibt sich die Architektur der Abbildung 4.1.3–3. Der Vorteil dieser Darstellung gegenüber dem Original von McClelland & Rumelhart ist, dass die gesamte Anordnung abgebildet werden kann, nicht nur Teile (noch dazu in einer leicht misszuverstehenden Form!).

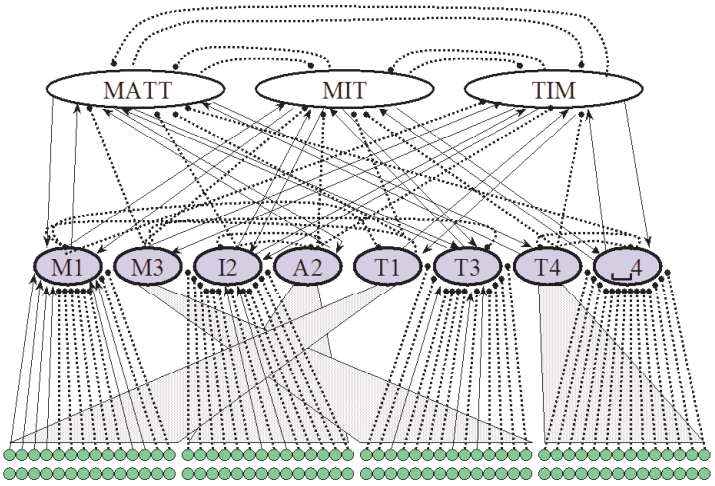

Abbildung 4.1.3–3: „Interactive activation"-Modell für die Wörter *matt, mit* und *Tim.* Die Verbindungen von den Merkmalseinheiten zu den Buchstabeneinheiten sind aus darstellungstechnischen Gründen teilweise nur durch grau eingefärbte Flächen wiedergegeben. Mit derselben Begründung sind die Verbindungen von der unteren Reihe der Merkmalseinheiten zu den Buchstabeneinheiten weggelassen. Ihr Verbindungsmuster ist genau komplementär zu dem der oberen Reihe: Wenn eine Einheit in einer bestimmten Position in der oberen Reihe eine erregende Verbindung zu einer Buchstabeneinheit hat, hat die entsprechende Einheit der unteren Reihe eine hemmende Verbindung.

Der Input kann als Repräsentation von zwei Bitmustern aufgefaßt werden, bei denen jeweils eine Null für ein nicht aktiviertes und eine Eins für eine aktiviertes Inputelement steht. Das eine der Bitmuster drückt das Vorhandensein der entsprechenden Merkmale aus, das andere Bitmuster das Fehlen

der entsprechenden Merkmale. Der Buchstabe M wird danach beschrieben durch die Bitmuster

11110000001100 und
00001111110011.

Die Nullen bedeuten also nicht etwa hemmende Aktivationen, sodass eine störungsfreie Eingabe des M zu insgesamt 14 erregten Eingängen in die Buchstabeneinheit führt, während keine wirksamen Hemmungen ausgelöst werden.

Abweichend von der Originalfassung des IA-Modells ist in der Abbildung 4.1.3–3 der Fall unterschiedlicher Wortlängen aufgenommen. Die doppelte Repräsentation des Inputs ist eine unverzichtbare Voraussetzung, wenn dieser Fall ohne künstliche Zusatzannahmen gelöst werden soll.

Zwei die Architektur betreffende Punkte sind besonders bemerkenswert:

- Die Einheiten der Buchstabenebene und die Einheiten der Wortebene sind jeweils hemmend untereinander verknüpft, das heißt, es entsteht ein Wettbewerbsverhalten, das im Idealfall nach einigen Simulationszyklen zur Auswahl eines Elements als Sieger führt (vgl. Teil 2, „Grundlagen“, Kap. 2.3.2).
- Es gibt erregende Rückverbindungen von der Wortebene zur Buchstabenebene, so dass die Aktivation einer Worteinheit die Aktivation entsprechender Buchstabeneinheiten verstärken kann.

Die bekannteste Leistung des IA-Modells besteht in der „Restaurierung“ defekter oder teilweise verdeckter Buchstaben. Das immer wieder besprochene und abgebildete Beispiel ist das Wort WORK, bei dem das K durch einen Tintenklecks so verdeckt ist, dass unklar wird, ob es sich um ein K oder ein R handelt. In der Simulation wird diese Situation dadurch abgebildet, dass in dem Paar von Eingabemustern alle Merkmale, die K und R differenzieren, eine Null haben, also als „off“ gelten. Die Simulation entscheidet (aufgrund der Wettbewerbsvorgänge und der Rückverbindungen von der Wortebene auf die Buchstabenebene) für K, weil das Lexikon nur WORK, aber nicht WORR enthält, das heißt, die Buchstabeneinheit K für die vierte Position im Wort hat gegenüber R eine deutlich erhöhte Aktivation, obwohl K und R im Input gleichberechtigt waren.

Die Architektur des Modells ist gekennzeichnet durch einen großen Aufwand an Einheiten im Inputbereich verglichen mit der Sparsamkeit in höheren Bereichen der Verarbeitung. Die Verhältnisse relativieren sich zwar, wenn mehr

Wörter aufgenommen werden, es bleibt aber ungewöhnlich, dass Buchstabeneinheiten pro Wortposition repräsentiert werden. Intuitiv würde man erwarten, dass es so etwas wie das Muster des Buchstabens M gibt, das dann in verschiedenen Positionen in verschiedenen Wörtern verwendet wird. Das würde dem Schema **A** in Abbildung 4.1.1–5 entsprechen. Die Annahme des IA-Modells entspricht aber Schema **B**. Das heißt, das Muster existiert nur in der Form ortsspezifischer Exemplare. Die Entscheidung für eines der Exemplare muss zunächst unabhängig vom Verlauf des Lexikonabgleichs geschehen.

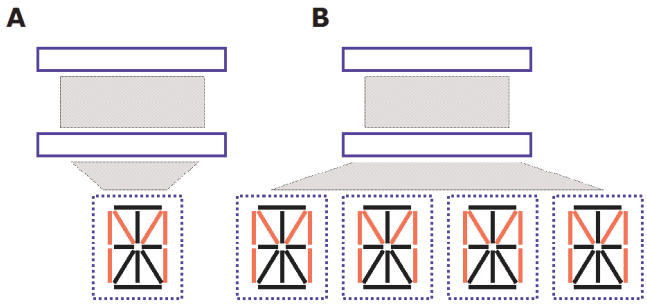

Abbildung 4.1.3–4: Wünschenswerte (**A**) und tatsächliche Struktur (**B**) des Systems.

IA modelliert Geschriebenes. Die Alphabetschrift bildet die phonologische Reihenfolge in der Zeit räumlich ab. Daher ist das Verfahren für ein Lesemodell vielleicht noch akzeptierbar. Schwieriger wird es bei der Übertragung des Verfahrens auf einen akustischen, zeitlich linearen Input. Das ist in dem Modell TRACE (McClelland & Elman, 1986) konsequent durchgeführt. Ordnungskriterium für Sequenzelemente ist jetzt nicht die (räumliche) Position im Wort, sondern die Position in der Zeit. Das auf die Zeit bezogene Reihenfolgekriterium hat zur Folge, dass jetzt auf allen Ebenen, auch auf der Wortebene, Exemplare pro Zeiteinheit erforderlich sind, also z. B. so viele Einheiten für ein und dieselbe Wortform, wie Zeittakte vorgesehen sind. (Aus praktischen Gründen wird innerhalb des Modells teilweise davon abgewichen.) Es ist natürlich viel schwieriger, eine Anzahl für die zu betrachtenden Zeittakte festzulegen, als eine Anzahl für die zu erwartenden Phonemeinheiten in einem Wort. Alle anderen wesentlichen Charakteristika des IA-Modells werden beibehalten: Das Modell verwendet durchgehend lokalistische Repräsentationen. Es gibt drei Ebenen von Einheiten, eine Ebene phonetischer Merkmale, eine Phonemebene und eine Wortebene. Zwischen

den Ebenen werden bidirektionale erregende Verbindungen angenommen, innerhalb der Ebenen sind konkurrierende Einheiten bidirektional hemmend verbunden.

Insgesamt entsteht dadurch das für diesen Modelltyp charakteristische Wettbewerbsverhalten, das sich wesentlich von dem Verhalten des frühen(!) Kohortenmodells von Marslen-Wilson unterscheidet, obwohl das „kohortentypische“ Abnehmen der Menge herausgehobener Kandidaten im Verlauf der Wortwahrnehmung erhalten bleibt. Von McClelland & Elman (1986: 99 ff.) werden zwei Kritikpunkte bezüglich des ursprünglichen Kohortenmodells besonders betont: Schwierigkeiten bei Störungen des Wortanfangs, die im ursprünglichen Kohortenmodell die Bildung einer nicht-leeren Anfangskohorte verhindern würden, und, damit verwandt, Schwierigkeiten, die durch Unsicherheiten bei der Wortsegmentierung entstehen. Beide Probleme werden durch TRACE gelöst. Wenn man genauer hinsieht allerdings mit einem unerwünschten Nebeneffekt, der darin besteht, dass zwar am Ende plausible Wortkandidaten durch hohe Aktivitäten herausgehoben werden, es bleibt aber offen, *wann* der Entscheidungsprozess ein brauchbares Ende gefunden hat. In den in der Originalarbeit referierten Beispielen ist das oft erst einige Phonemdauern nach dem Wortende der Fall. Das dürfte der alltäglichen Beobachtung, dass man ein Wort oft erkennen kann, ehe es verklungen ist (die dem Kohortenmodell ursprünglich zugrundegelegen hat) deutlich widersprechen.

Wichtige weitere Mängel des TRACE-Modells werden von den Autoren selbst gesehen (119 f.):

- TRACE kann nicht durch Lernvorgänge entstehen.
- Die große Zahl von Wiederholungen gleichbedeutender Einheiten (50 Lexika?) ist unplausibel.
- Es gibt Schwierigkeiten mit Variationen der Sprechgeschwindigkeit und individuellen oder kontextabhängigen phonetischen Parametern.
- Es gibt Schwierigkeiten mit der Modellierung von Priming-Effekten.

Wir übergehen an dieser Stelle einige bekannte Modelle, die mit Triphon-Repräsentationen (siehe oben 4.1.1, in konnektionistischer Literatur meist als „Wickelphone“ bezeichnet) zur Darstellung lexikalischer Sequenzen arbeiten, wie Rumelhart & McClelland (1986b) und Seidenberg & McClelland (1989). In diesen Modellen werden eher Teilfunktionen des Lexikons dargestellt (Zuordnung von Vergangenheitsformen, Zusammenhang zwischen orthographischer und phonetischer Repräsentation) als dass es wirklich um

Lexikonstrukturen im Allgemeinen geht. Die Entwicklung der 90er Jahre zeigt, dass moderne konnektionistische Lösungen eher zu rekurrenten Netzen führen. Rekurrente Netze sind dadurch charakterisiert, dass sie die zeitliche Sequenz als zeitliche Sequenz im Modell beibehalten. Es wird zum Beispiel ein Sequenzelement innerhalb eines Zeittakts repräsentiert, das folgende Element im folgenden Zeittakt usw. Der Zustand zu einem bestimmten Zeittakt beeinflusst den Zustand im folgenden Zeittakt. Diese Vorhersagefunktion wird meist dadurch realisiert, dass von der Schicht der „hidden units“ ausgehend eine Zustandskopie gebildet wird, die in den Berechnungsvorgang des folgenden Zeittakts einbezogen wird. In Abschnitt 2.2.3 des Teils 2 ist in Abbildung 2.2.3–2 ein solches rekurrentes Netz zusammen mit einem zugehörigen Beispiel wiedergegeben.

Norris (1990) kann an einem vereinfachten Modell zeigen, dass rekurrente Netze gegenüber unterschiedlichen Eingabegeschwindigkeiten (Eingabeelemente werden über eine unterschiedliche Anzahl von Zeittakten wiederholt) gradmäßig unempfindlicher sind als Netze, die ausschließlich Vorwärtsverbindungen haben und den zeitlichen Ablauf in der Art des Modells TRACE repräsentieren. Die Wiedererkennungsleistung ist allerdings mit 80% auch nicht besonders befriedigend. Für die Modellierung realistischer Funktionen des Lexikonabgleichs wird das Netz der Abbildung 4.1.3–4 eingesetzt. Es wird unter Verwendung des Backpropagation-Algorithmus trainiert, ist also nicht, wie TRACE „hard wired“. Bemerkenswert ist, abgesehen von den für rekurrente Netze typischen Feedback-Verbindungen, die eine Zeitverzögerung von einem Zeittakt haben, dass die Ausgabeschicht lokalistisch interpretiert wird, das heißt, sie enthält 50 Einheiten zur Darstellung von 50 Wörtern. Die Eingabeschicht enthält 11 „binary speech features“ zur Repräsentation von Segmenten, die als Buchstaben oder Phoneme interpretiert werden können. Es wird *ein* in dieser Form repräsentiertes Segment pro Zeittakt eingegeben.

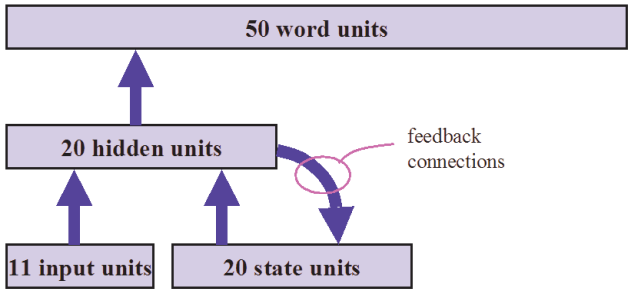

Abbildung 4.1.3–4: Rekurrentes Lexikonnetz bei Norris (1990).

Norris vergleicht die Leistungen seines Modells mit den Annahmen des Kohortenmodells von Marslen-Wilson und stellt eine weitgehende Übereinstimmung fest. Hohe Aktivationswerte der Ausgabeeinheiten ergeben sich unmittelbar mit dem Erreichen der Eindeutigkeit der Eingabekette. Abweichungen, die das Gewicht der Wortanfänge für das Abgleichsergebnis betreffen, werden von Norris positiv bewertet: Das Netz ist toleranter gegenüber kleineren Störungen der Wortanfänge, als in der strengen Interpretation des Kohorten-Modells vorgesehen.

Dieses rekurrente Lexikonmodell ist in Norris (1994) erweitert worden zu einem als SHORTLIST bezeichneten Modell, das allerdings insofern ein Kuriosum darstellt, als fast alle wesentlichen Teile des Modells nicht in konnektionistischer Architektur ausformuliert, sondern in klassischem Programmierstil hergestellt worden sind. Wenn man die Komponenten doch konnektionistisch auffasst, besteht SHORTLIST aus einem rekurrenten Teilnetz, das dem Netz aus Norris (1990) entspricht, und einem Wettbewerbsteil der üblichen Form, in dem konkurrierende Wortknoten jeweils inhibitorisch verknüpft sind. Die konkurrierenden Wortknoten des Wettbewerbsnetzes werden aus dem Ergebnis des rekurrenten Netzes abgeleitet. Wenn es zu viele sind, wird die Liste nach dem Kriterium der Aktivationsstärke verkürzt. Das Wettbewerbsnetz ist, entsprechend dem laufenden Input in das rekurrente Netz, der ja Phonem für Phonem erfolgt, laufenden Veränderungen unterworfen:

> „As each new phoneme is presented to the model the lexical search procedure first updates the candidate sets and rewires the network as necessary. The lexical search also updates the bottom-up activation score for each candidate. The lexical competition network then cycles through a fixed number of cycles (15 cycles in all of the present simulations) before the next phoneme is presented.“ (Norris, 1994: 204 f.)

Ein alternatives rekurrentes Modell wird von Gaskell & Marslen-Wilson (1997) vorgeschlagen. Es verwendet ausschließlich verteilte Repräsentationen, die Ausgabeschicht ist zweigeteilt, in „lexical semantics“ (binäres Muster über 50 Einheiten) und „phonology“. Der Input besteht in 11 binär kodierten phonetischen Merkmalen. Während des Inputverlaufs verändert sich der Abstand der verteilten Outputrepräsentationen zu den Zielwerten in der von der Kohortentheorie vorhergesagten Weise. Auch hier wird, wie bei rekurrenten Netzen üblich, zum Training der Backpropagation-Algorithmus verwendet.

4.1.4 Zusammenfassende Bewertung

Dass irgendeines der vorgeschlagenen Modelle mit den biologischen Bedingungen der Sprachverwendung oder auch nur mit akzeptierten alltäglichen Beobachtungen vollständig vereinbar wäre, wird niemand behaupten wollen. Die Frage kann also nur sein, wie viele offenkundig unbiologische und intuitiv problematische Begleitannahmen man akzeptieren möchte.

Das Lexikon gilt in der Linguistik als derjenige Teil der sprachlichen Kompetenz, der am wenigsten universell ist, der also in jedem Fall durch Lernvorgänge im Zuge des Spracherwerbs aufgebaut werden muss. Es ist verwunderlich, dass die meisten Lexikonmodelle darauf verzichten, die erforderlichen Lernprozesse überhaupt in die Betrachtung einzubeziehen. Das Thema „Lernen“ wird eigentlich erst im Zusammenhang mit rekurrenten konnektionistischen Netzen ernst genommen. Das geschieht aber dort durch Anwendung des Backpropagation-Algorithmus, das heißt eines Lernalgorithmus, von dem alle Welt weiß und akzeptiert, dass er biologisch unrealistisch ist (vgl. Teil 2, „Grundlagen“ Abschnitt 2.3.2).

Es ist selbstverständlich, dass Details der Implementierung in den Modellen der Computerlinguistik und der künstlichen Intelligenz nicht wörtlich auf die natürliche Sprachverarbeitung im Menschen übertragen werden können. Die damit entstehende Lücke wird aber in den neuro- und psycholinguistischen Modellen nicht ausgefüllt. Es ist charakteristisch, dass Implementierungsdetails mehr oder weniger ausdrücklich ausgeklammert werden, es werden also Vorgänge angenommen, von denen völlig unbekannt bleibt, durch welche Art von Apparatur sie zustande gebracht werden könnten. Wie soll, um ein Beispiel zu geben, eine Worteinheit mit den von Morton (1969) oder Marslen-Wilson (1980 und 1987) verlangten Eigenschaften konkret aussehen? Der Übergang der Kohortentheorie in das konnektionistische Lager ist unmittelbar einzusehen.

Meist ist nicht abzuschätzen, wie die in den Modellen verwendeten Vorgänge in der beobachtbaren Verarbeitungszeit unterkommen. Wenn es, naturgemäß hauptsächlich bei konnektionistischen Modellen, doch Anhaltspunkte gibt, sind die vom Modell geforderten Verarbeitungszeiten durchweg unrealistisch. Wenn man die Angaben von Norris (1994) zu SHORTLIST in dem oben zitierten Text ernst nimmt, müsste ein an dem Wettbewerbsprozess beteiligtes Neuron mindestens 15 Aktionspotenziale (ein Aktionspotenzial pro Zeitzyklus) abgeben für jedes verarbeitete Phonem. Wenn erreicht werden soll, dass 15 Phoneme pro Sekunde bewältigt werden, was einer durchaus nicht überhasteten Sprechweise entspricht, muss ein solches Neuron 225 Aktionspotenziale produzieren, das ist eine um den Faktur 4 höhere Frequenz,

als für den Kortex im Allgemeinen von Biologen, die an hohen Frequenzen interessiert sind, angenommen wird. Die von SHORTLIST vorausgesetzten Vorgänge sind also an der Realität gemessen viel zu langsam.

Ähnlich unklar wie der Input in den Lexikonabgleich ist der Wert des Outputs. Ein Problem ist die Festlegung des Zeitpunkts, zu dem ein gültiger Output erzielt ist. Bei der Kohortentheorie in der ursprünglichen Version war der Anhaltspunkt die Kohortengröße. Eine auf ein einzelnes Element geschrumpfte Kohorte war ein gültiges Verarbeitungsergebnis und konnte sofort weiterverarbeitet werden. Wenn man allerdings mit Nicht-Wörtern rechnen muss, das heißt damit, dass das letzte verbliebene Element sich ebenfalls noch in kohortenüblicher Weise ausschalten kann, reicht dieses Kriterium nicht aus. Es reicht auch nicht aus, wenn man nach Marslen-Wilson (1987) mit gradmäßiger statt binärer Zuordnung zur Kohorte rechnet. Es entstehen dann ähnliche Probleme wie bei dem Modell TRACE, bei dem die Wartezeit bis zum Erreichen eines gültigen Outputs unbestimmt bleibt.

Bei Modellen mit verteilten Repräsentationen, wie dem von Gaskell & Marslen-Wilson (1997), man vgl. auch Gaskell & Marslen-Wilson (2002) und Rodd, Gaskell & Marslen-Wilson (2004), ist die Qualität des Outputs besonders bei Mehrdeutigkeit schwer einzuschätzen, und die Konsequenzen für die Weiterverarbeitung bleiben unklar.

Schließlich soll wenigstens mit einem Satz darauf hingewiesen werden, dass alle in Teil 2, „Grundlagen“, gegen symbolverarbeitende und konnektionistische Modelle (auch abgesehen von der Problematik der Lernalgorithmen) angeführten Argumente zusätzlich gelten.

Die Beschäftigung mit den existierenden Modellen ist selbstverständlich trotz aller Mängel nicht sinnlos. Was bleibt, sind Bündel nützlicher Beobachtungen. Es handelt sich um Beobachtungen wiederkehrender Eigenschaften an Modellen und Beobachtungen bezüglich einzelner Leistungen, die für oder gegen die verschiedenen Modelle ins Feld geführt worden sind. Hier einige Beispiele:

Der frühe „recognition point“ in Wortentscheidungsaufgaben ist ein wichtiger Anhaltspunkt. Es folgt daraus zunächst, dass Wörter nicht als Ganzes im Lexikon gesucht werden, sondern der Abgleich kann offenbar Segment für Segment erfolgen, jedenfalls bei auditivem Input. In diesem Fall ist dann auch die Unterschiedlichkeit der Wortlängen in den natürlichen Sprachen kein kritischer Faktor. Der Lexikonabgleich muss aber, wenn man die Lexikonstrukturen der computerlinguistischen Modelle und ihre Folgen betrachtet, nicht buchstäblich so verstanden werden, dass eine Sammlung (Kohorte) von Wörtern gebildet und dann sukzessive reduziert wird, das

Zusammenlegen der Wortanfänge in den Strukturen der Ausdrucksseiten hat analoge Konsequenzen.

Es ist vermutlich nicht notwendig, dass alle in einem lexikalischen Ausdruck verborgenen lexikalischen Teilketten gesucht und als Kandidaten bearbeitet werden. Es ist sinnlos, bei dem Input /lampe/ auch die Präposition /am/ zu identifizieren, wie das in konnektionistischen Modellen in der Regel vorgesehen ist. Stattdessen kann ein „definierter Lexikonstart" angenommen werden, der erst dann eine neue Kette erwartet, wenn mindestens *eine* laufende Kette einen Abschluss gefunden hat. (Hinweise zu diesem Thema bei Frauenfelder & Floccia, 1999: 30 ff.)

Mit einer Mehrdeutigkeit lexikalischer Ketten im Verstehensprozess ist immer zu rechnen. Zeitparallele Verarbeitung ist angesichts der „Langsamkeit" elementarer biologischer Vorgänge im Kortex unverzichtbar. Es ist aber fraglich, ob Mehrdeutigkeit durch Unschärfe ersetzt werden kann, wie in verteilten konnektionistischen Systemen üblich. Es ergeben sich jedenfalls unterschiedliche Auswirkungen auf die Möglichkeiten der Weiterverarbeitung auf „höheren" Ebenen.

Es ist, wie SHORTLIST zeigt, nicht notwendig, selbst unter sonst standardmäßigen konnektionistischen Architekturvorstellungen und in rekurrenten Netzen, mit der Annahme verteilter Repräsentationen auf allen Repräsentationsebenen zu arbeiten.

Die zeitliche Variabilität des Sprachsignals ist besonders für die Sprachperzeption ein wichtiges zu lösendes Problem. Es bleibt zu klären, wie viel von dieser Variabilität im prälexikalischen Bereich abgefangen werden kann, und wieviel als Belastung für den eigentlichen Lexikonabgleich übrig bleibt.

Insgesamt bleibt festzuhalten, dass es nicht genügt, Lexikonmodelle daran zu messen, wie gut sie Daten aus psycholinguistischen Experimenten wiederspiegeln. Sie müssen auch zu den apparativen Voraussetzungen passen, die durch die biologische Grundlage gegeben sind.

4.2 Anforderungsprofil

4.2.1 Allgemeine Gesichtspunkte

Wenn man den Weg des auditiven Signals vom Ohr ausgehend verfolgt, dann bleibt dieses Signal bis in den Kortex hinein zunächst monomodal und kann vom linguistischen Standpunkt aus als „lautlich“ charakterisiert werden. Ab einem bestimmten Punkt, der schon im Kortex liegt, gilt diese monomodale Charakteristik nicht mehr. Man beachte, dass die Natur des Signals sich dabei nicht notwendig ändert, es ist die Vorgeschichte, die einen Unterschied ausmacht. Einem einzelnen Aktionspotenzial oder auch einer Folge von Aktionspotenzialen, auf einer einzelnen Zelle ausgelöst, kann nichts davon angemerkt werden. Vom linguistischen Standpunkt aus betrachtet ist der Übergang zur nicht mehr monomodalen Charakteristik identisch mit dem Übergang von sprachlichem Ausdruck zu sprachlichem Inhalt. Was als mentales sprachliches Lexikon angesehen wird, ist sozusagen um diesen Umschlagspunkt herum angesiedelt. (Es ist zu beachten, dass nicht nur visuelle, taktile, olfaktorische usw. Stimuli oder darauf basierende Gedächtnisinhalte als zu einer anderen Modalität gehörend zu betrachten sind, sondern auch auditive, außerhalb des *aktuellen* sprachlichen Signals selbst.)

Wenn man die Dinge so beschreibt, ist von vornherein klar, dass das mentale Lexikon nicht als Informationsmenge gesehen wird, die von einem Sprachprozessor (was immer man sich darunter vorstellen mag) verwendet wird, sondern dass der Unterschied zwischen Speicherung und Verarbeitung (Daten und Prozessen/Prozessoren) aufgehoben ist. Dass ist eine Grundbedingung für die gerade im lexikalischen Bereich, wie schon die psycholinguistischen Modelle zeigen, unverzichtbare Voraussetzung zeitparalleler Verarbeitung. Damit sind auch symbolverarbeitende Vorgänge, die eine Adressierung von Speicherelementen voraussetzen, ausgeschlossen.

Die Möglichkeit von Parallelverarbeitung im Kortex ist, nach den Überlegungen von Teil 2, „Grundlagen“, an eine lokalistische Architektur gebun-

den, da sonst nur biologisch unplausibles „überwachtes“ Lernen möglich wäre. Das mentale Lexikon unterliegt lebenslang, nicht nur im Primärspracherwerb, ständiger Veränderung, die nicht nur Vergessen, sondern auch Erweiterung und Anpassung einschließt. Es muss effektive „unüberwachte“ Lernprozesse geben, die zu solchen Veränderungen führen. Das lokalistische Repräsentationsprinzip ermöglicht es darüber hinaus, dass bei Mehrdeutigkeit im Sprachverstehensprozess tatsächlich mehrere unterschiedliche Verarbeitungspfade gleichzeitig verfolgt werden können. Die Entwicklung des Kohortenmodells (siehe oben Abschnitt 4.1.2) zeigt, dass offenbar Entscheidungen bei Mehrdeutigkeit möglichst lange offen gehalten werden sollten.

Wenn man Lernprozesse etwas genauer analysiert, stellt man fest, dass nicht alle erforderlichen Verbindungen zwischen Neuronen durch Lernprozesse etabliert sein können. Das gilt insbesondere für hemmende Verbindungen. Daraus ergibt sich allgemein, dass mit einer vererbten festen Verbindungsstruktur gerechnet werden muss, die im Kortex ja auch tatsächlich erkennbar ist, und die zu unterscheiden ist von einem ebenfalls vorhandenen Angebot variabler Verbindungen. Für das Angebot variabler Verbindungen gilt, dass es im Wesentlichen chaotisch sein muss.

Schließlich ist in Teil 2 auch gezeigt worden, dass im Kortex, in allen Bereichen, in denen Lernprozesse der beschriebenen Art möglich sein sollen, das Signal (wenn man überhaupt davon sprechen darf), das von einer Zelle zur anderen weitergeleitet wird, in einem einzelnen Alles-oder-nichts-Impuls besteht, so dass Erregungsintensitäten nicht von Zelle zu Zelle weitergegeben werden können. Wenn Aktivationstärke eine Rolle spielt, bleibt sie auf die einzelne Zelle beschränkt und ist durch das jeweilige „exzitatorische postsynaptische Potenzial“ (EPSP; vgl. Teil 2, Kapitel 2.1.3) realisiert.

Wenn man die im vorigen Kapitel vorgestellten Lexikonmodelle in der Psycholinguistik und im Konnektionismus daraufhin untersucht, wie sie mit anderen Komponenten der Sprachverarbeitung verknüpft zu denken sind, stellt man fest, dass das in den meisten Fällen nicht thematisiert wird. Eine der Ausnahmen ist das Sprachproduktionsmodell von Levelt (siehe oben Abschnitt 4.1.2). Es kann aber nicht darum gehen, sich Lexikonstrukturen vorzustellen, die jeweils mit einigen Laborexperimenten übereinstimmen und an die keine weiteren Ansprüche gestellt werden. Es ist von vornherein zu fordern, dass eine konstruierte Lexikonstruktur einbettbar ist in alle betroffenen „umgebenden“ Prozesse. Im Zusammenhang damit muss beachtet werden, dass die Reaktionen von Versuchspersonen nicht leichtfertig als direkte Folge lexikalischer Eigenschaften interpretiert werden sollten. Sie dürften in der Regel an komplexe Prozesse gebunden sein, die nicht nur in der Einleitung motorischer Vorgänge bestehen.

4.2.2 Prälexikalische Repräsentationen

Es geht in diesem Abschnitt um die Abgrenzung eines Bereichs, den man als lexikalisch bezeichnen möchte „nach unten“, und um die Charakterisierung der Informationen, die an dieser Schnittstelle anstehen. Die Perspektive ist die der Sprachperzeption, nicht der Produktion.

Frauenfelder & Floccia (1999: 5) weisen darauf hin, dass einige Probleme in dieser Hinsicht durch die Arbeitsteilung in Phonetik/Phonologie einerseits und Psycholinguistik (des Lexikons) andererseits entstehen:

> „Ein wenig zufriedenstellendes Ergebnis dieser Arbeitsteilung besteht darin, dass einige Hauptprobleme beim Erkennen von Wörtern und gesprochener Sprache nur zu leicht zwischen die Interessen der beiden Disziplinen fielen und unberücksichtigt blieben. Beispielsweise bleibt ungelöst, auf welcher Ebene – auf prälexikalischer oder lexikalischer Ebene oder auf beiden – der Hörer das Problem der dem Sprechsignal innewohnenden Variabilität handhabt und wo der Forscher dementsprechend ansetzen sollte. Ähnliche Mehrdeutigkeiten ergeben sich bei der Segmentierung des kontinuierlichen gesprochenen Inputs: Findet der Übergang von kontinuierlichen zu diskreten Kategorien auf phonetischer Ebene, auf lexikalischer Ebene oder auf beiden Ebenen statt?“

Was man in beiden Arbeitsbereichen, sowohl in der Phonologie, als auch in der Psycholinguistik des Lexikons antrifft, ist die Rede von Silben, Halbsilben (Dupoux, 1993), oder Moren. Da es sich dabei in jedem Fall um prinzipiell zusammengesetzte Gebilde, das heißt spezieller: zeitliche Ketten von noch kleineren Einheiten handelt, entstehen dadurch ähnliche Fragestellungen und Probleme wie bei der Behandlung von Lexikoneinträgen (als Phonemketten verstanden), so dass wir diese Gebilde aus technischen Gründen dem lexikalischen Bereich zuordnen. Die Silbenproblematik wird konsequenterweise vertagt, bis etwas genauere Vorstellungen über grundlegende lexikalische Strukturen und Prozesse vorliegen.

Wenn Einheiten wie die Silbe zum lexikalischen Bereich gerechnet werden, bleibt die Phonemebene als mögliche prälexikalische Grenzstruktur. Aus den Überlegungen in Teil 3, „Phonetik/Phonologie“, ergibt sich, dass mit Phonemen als Einheiten tatsächlich zu rechnen ist. Es ist allerdings nicht hilfreich, sich vorzustellen, dass Phonemketten, die dann als Basis des Lexikonabgleichs dienen sollen, in einer Art „Buffer“ zwischengespeichert werden (siehe unten 4.2.3). Phoneme sind Einheiten, die zu je bestimmten Zeitpunkten dem Prozess des in der Zeit sich entwickelnden Lexikonabgleichs

logisch vorgeschaltet sind, wobei ein Lexikoneintrag die Repräsentation einer zeitlichen Folge solcher Einheiten darstellt.

An dieser Stelle sollte die Thematik der Phonemebene aber auch noch einmal mit Bezug auf das Lexikon etwas genauer analysiert werden. In den jüngeren Versionen des Kohortenmodells wird ausdrücklich auf eine Phonemebene verzichtet, und der Input in das Lexikonmodell von Gaskell & Marslen-Wilson (1997) besteht direkt in einem Muster von phonetischen Merkmalen.

Das Extrem einer solchen Strategie wäre, dass das Ergebnis der Frequenzanalyse der Cochlea, also z. B. bei Vokalen das Muster der Formanten, direkt als Input in den Abgleichsprozess dient. Das ist allerdings nur dann möglich, wenn man annimmt, dass die entsprechenden Informationen im Kortex, angesichts der niedrigen im Kortex beobachteten Impulsfrequenzen, überhaupt ankommen. Ein anderes, etwas komplizierteres Argument ist, dass es Phonemdifferenzen gibt, die in einer Sprache verwendet werden, aber von Sprechern einer anderen Sprache nicht wahrgenommen werden, was einen Vergessensprozess voraussetzt, wenn man annimmt, dass dieselben Differenzen bis zu einem Alter von sechs Monaten noch erkannt werden können. Wenn die Auswertung des Lautspektrums erst eine Angelegenheit des Lexikons wäre, müsste es leicht sein, solche Phonemdifferenzen neu zu lernen, da doch das Nachsprechen und Erinnern unbekannter Wörter im Rahmen des Lautinventars einer Sprache problemlos ist, also entsprechende Lernprozesse im lexikalischen Bereich sehr effektiv sein müssen.

Das ist aber nicht der Fall. Also wird man die Auswertung des Lautspektrums Verarbeitungsprozessen und Strukturen zuweisen, die unterhalb des Lexikons liegen. Das Ergebnis dieser Verarbeitungsprozesse kann dann als eine Kombination phonetischer Merkmale betrachtet werden (ggf. auf angeborenen Grundlagen beruhend), die dann, den Vostellungen von Gaskell & Marslen-Wilson (1997) entsprechend, Input in den Lexikonabgleich sein könnten. Hier ist aber zu beachten, dass für den Kortex gelten muss, dass überall rasche Konzeptbildungsprozesse möglich sind, die zu definierten Großmutterzellen führen. Jedenfalls ist es bei der großen Homogenität der Kortexstruktur nicht einfach, davon auszugehen, dass gerade im auditiven Bereich diese Regel nicht gilt. Da die Merkmale, die aus einem Lautspektrum abgeleitet werden, innerhalb eines kleinen Zeitfensters von vielleicht 30ms aktiviert sein müssen, also praktisch „gleichzeitig“ Impulse abgeben, ist ein Konzeptbildungsvorgang unvermeidbar. Damit gilt, dass Phoneme als Einheiten zwangsweise entstehen (Phoneme können nicht angeboren sein), es ist allerdings dadurch noch nicht gesagt, dass es sich um die Phoneme handelt, die üblicherweise in der Linguistik unterschieden werden. Für Phoneme gilt

damit die gleiche Variabilität wie für andere Konzepte auch, einschließlich der möglichen Prototypizität, vgl. Teil 2, „Grundlagen“, Kapitel 2.4.5.

Es ist also damit zu rechnen, dass es Phoneme sind, die den Input in den Lexikonabgleichsprozess bilden. Ein Phonem wird als Konzept lokalistisch durch eine Großmutterzelle repräsentiert (genauer: redundant durch eine Menge etwas komplexerer gleichwertiger Großmuttereinheiten). Wenn ein Phonem in einer Lautkette vorhanden ist, gibt diese Großmutterzelle mindestens ein einzelnes Aktionspotenzial ab, bei größerer zeitlicher Länge auch mehrere. Phoneme können neuronal im Kortex nicht so dargestellt werden, dass die Länge auch des kleinsten vorkommenden Segments durch entsprechend ausgedehnte Impulszüge abgebildet wird. Dazu sind die beobachtbaren Impulsfrequenzen zu niedrig. Man muss außerdem das von Klatt (siehe oben 4.1.1) festgestellte Paradox beachten, nach dem Segmentgrenzen eigentlich erst festgestellt werden können, wenn man weiß, um welches Segment es sich handelt, was aber wiederum von der Feststellung der Segmentgrenzen abhängt.

Wenn aber Phoneme ihre zeitliche Erstreckung im Kortex verlieren, ist zu erwarten, dass ein gewisses Ausmaß an zeitlicher Variabilität einer Phonemsequenz vom Standpunkt des Lexikons aus gesehen problemlos verkraftet werden kann, entsprechende lexikalische Verarbeitungsstrukturen vorausgesetzt.

Ein phonologischer Input der beschriebenen Art löst bis zu einem gewissen Grad sowohl das Problem der zeitlichen als auch das der qualitativen Variabilität des Sprachsignals. Zusätzlich gilt, dass ein Phonemkonzept, in einer spezialisierten Zelle repräsentiert, nur dann einen Impuls abgeben kann, wenn es erkannt ist, das heißt, wenn die Integration der Inputdaten die phonemspezifische Schwelle erreicht hat. Es können deshalb nicht beliebige Segmente (als zeitliche Erstreckungen interpretiert) in das Sprachsignal hineingelegt werden. Die Integrationspunkte bilden das gültige zeitliche Raster.

Mehrdeutigkeit schließlich bedeutet im prälexikalischen Bereich, dass verschiedene Phoneme gleichzeitig aktiviert werden. Mehrdeutigkeit kann dazu führen, dass mehrere lexikalische Ketten gleichzeitig abgearbeitet werden und nur diese, nicht zusätzliche, eventuell konkurrierende, aber lautlich abweichende. Bei zwei Alternativen ist die Unterscheidung leichter, wenn sie über beide Ohren getrennt wahrgenommen werden.

Tonbeispiele:
Beispiel 1: *bieten* links
Beispiel 2: *booten* rechts
Beispiel 3: *bieten+booten* stereo
Beispiel 4: *bieten+booten* mono

Auf diese Weise können Entscheidungen über das Vorliegen einer Wortform (wahrscheinlich ausreichend weit) aufgeschoben werden, was letzlich eine selbstverständliche Leistung eines beliebigen parallelverarbeitenden lokalistischen Systems ist.

Bleibt die ergänzende Feststellung, dass Phoneme nicht nur bequeme Abkürzungen für die eigentlich allein existierenden Merkmalsbündel sind, was dann auch die Konsequenz hat, dass die Annahme redundanzfreier lexikalischer Repräsentationen nach Art der älteren generativen Phonologie (Chomsky & Halle, 1968) unmöglich wird.

Einige ganz wesentliche Lasten des lexikalischen Verarbeitungsprozesses werden, wie man sieht, durch die Phonologie abgefangen. Das gilt nicht für die Segmentierung der Eingabekette oberhalb der Phonemebene, also die Segmentierung in Wortbestandteile oder Wörter, die Sequenzen von Phonemen sind; und unklar muss an dieser Stelle auch die Rolle der Prosodie bleiben.

4.2.3 Anforderungen an lexikalische Strukturen (Ausdrucksseiten)

Lexikalische Ausdrucksseiten bestehen, wenn man die Einheit des Phonems akzeptiert, aus zeitlichen Sequenzen von Phonemen. Das gilt jedenfalls für den natürlichsten Fall des Sprechens über die lautliche Artikulation und das auditive Verstehen. Lesen und Schreiben sind hier zunächst auszuklammern. Wenn man die Phoneme als Bündel von Komponenten betrachtet und die lexikalischen Ausdrucksseiten ebenfalls als komponentiell, wobei jetzt Phoneme die Komponenten sind, wird deutlich, dass im Rahmen der Sprachverarbeitung zwei Formen von Komponentialität unterschieden werden müssen: Im Fall der Komponentialität von Phonemen gelten die Komponenten als psychisch „gleichzeitig“, auch dann, wenn z. B. Verschluss und Öffnung physikalisch gesehen aufeinander folgen. Im Fall lexikalischer Sequenzen sind die Komponenten unterschiedlichen Zeitpunkten zugeordnet und die Abfolge ist über Lernvorgänge festgelegt. Es ist zu erwarten, dass dieser Differenz unterschiedliche neuronale Grundlagen entspechen.

Das Raster der Zeitpunkte, das die Sequenz ausmacht, ist flexibel und in gewissen Grenzen charakteristischen Schwankungen unterworfen. Obwohl ein Teil der zeitlichen Variabilität im Verstehensprozess auf der Phonemebene lokalisiert bleibt, ist es doch so, dass man, wenn man im Verlauf des Sprachsignals pro Phonem ein Aktionspotenzial annimmt, diese Aktionspotenziale unterschiedliche zeitliche Abstände zueinander haben können.

Es ist oben in Abschnitt 4.2.2 schon angesprochen worden, dass man nicht mit einem „Buffer“ für phonologische Ketten rechnen kann. Die psycholinguistischen Experimente, die zur Entwicklung der Kohortentheorie (siehe oben 4.1.2) geführt haben, schließen aus, dass im Verstehensprozess Schallereignisse, die lexikalischen Ausdrucksseiten entsprechen, zwischengespeichert werden, ehe der Lexikonabgleich erfolgt. Ein solcher Vorgang würde wohl auch Adressierungsvorgänge im Sinne symbolverarbeitender Architekturen voraussetzen. Das bedeutet, dass sozusagen die Zeit, als zeitlich erstreckte Folge, nicht schon vor Erreichen der Lexikonrepräsentationen ausgewertet wird, sondern fortlaufend im Vorgang des Lexikonabgleichs. Die Repräsentation lexikalischer Ausdrucksseiten muss dem Rechnung tragen.

Das Problem der Sequenzdarstellung wird durch die oben in Kapitel 4.1 besprochenen Modelle nicht befriedigend gelöst. Das gilt insbesondere für die zeitliche Variabilität, für die erforderlichen Lernprozesse und bezüglich des ungeklärten Verhältnisses der in den konnektionistischen Modellen verwendeten Elementareinheiten zur tatsächlichen neuronalen Ausstattung des Gehirns. Zusätzliche Schwierigkeiten entstehen, wenn man versucht, Silben und andere, selbst schon sequenzielle Einheiten, als Elemente sequenzieller lexikalischer Repräsentationen zuzulassen.

Es ist möglich, sich die Menge der lexikalischen Ausdrucksseiten als listenförmig repräsentiert vorzustellen. Wohlgemerkt: nicht als eine Liste, die seriell durchsucht wird, wohl aber so, dass die einzelnen Ausdrucksseiten unabhängig voneinander bleiben und keine durch „Verklebungen“ mit anderen Ausdrucksseiten (siehe oben 4.1.1) gemeinsamen Bestandteile enthalten. Das entgegengesetzte Extrem wäre eine redundanzfreie Repräsentation, bei der alle gleichen Bestandteile der Ketten auch nur einfach vorhanden sind. Damit ist nicht nur die Frage gestellt, ob eine solche redundanzfreie Repräsentation überhaupt wünschenswert ist, sondern auch, ob sie aufgrund der anzunehmenden Lernprozesse entsteht und ob sie die erforderlichen Leistungen in den Verarbeitungsprozessen gewährleisten kann. Dass eine redundanzfreie Repräsentation auf phonologischer Ebene nicht möglich ist, ist schon oben 4.2.1 erläutert worden.

Mit dem Thema Redundanz oberflächlich verwandt ist auch die Frage, ob das mentale Lexikon ein Vollformenlexikon ist, also flektierte Wortformen als Ganzes repräsentiert sind oder ob auch gebundene Morpheme wie Flexionsendungen als einzelne, selbständige Einträge möglich sind. Der Trend in der diesbezüglichen Literatur geht dahin, dass man mit einer Mischung der

beiden Möglichkeiten rechnet, abhängig von den Erwerbsbedingungen (vgl. Sonnenstuhl-Henning, 2003).

Insgesamt ist also mit der Entscheidung für ein lokalistisches Repräsentationsprinzip und einer damit verträglichen Sequenzdarstellung das Problem des ausdrucksseitigen Lexikons nicht erschöpfend behandelt. Es muss nach einer übergeordneten Gesamtstruktur gefragt werden, die – bei der Annahme der Nichttrennung von Speicherung und Verarbeitung – die erforderlichen Leistungen erbringt und vor allem nach einer Struktur, die auch im Hinblick auf die lexikalischen Inhaltsseiten schlüssig ist.

4.2.4 Zusammenhang von Perzeptions- und Produktionsstrukturen

Engelkamp & Rummer (1999) sind Vertreter einer strikten Unterscheidung von Perzeptions- und Produktionssystem im lexikalischen Bereich, und zwar speziell im Bereich der lexikalischen Ausdrucksseiten. Die lexikalischen Ausdrucksseiten für die Sprachperzeption werden als „Wortmarken“ bezeichnet, die produktionsseitigen Äquivalente als „Sprechprogramme“ (wir übergehen die von Engelkamp & Rummer ebenfalls angenommenen getrennten Systeme für Schreiben und Lesen).

Die für diese Unterscheidung maßgebenden Argumente stammen aus Beobachtungen spezifischer aphasischer Störungen. Es sind Patienten beschrieben worden, deren Verstehensprozesse ungestört erscheinen, die aber unfähig sind, Wörter zu äußern. Der komplementäre Fall (im Sinne einer „doppelten Dissoziation“) sind Patienten, deren Spontansprache in Ordnung ist, die aber trotz intakten Hörvermögens (Messung mit reinen Sinustönen) Probleme mit dem Verstehen von Äußerungen haben (Worttaubheit).

Den unterschiedlichen ausdrucksseitigen Systemen steht ein gemeinsames konzeptuelles System gegenüber (Abbildung 4.2.4–1). Die *Einheit* eines Lexikoneintrags ist damit erst auf der inhaltlichen Ebene manifest.

Eines der Probleme bei dieser Vorstellung entsteht durch die Monitorfunktion des Hörens (Wahrnehmens) der eigenen Sprachproduktion. Der Produktionsprozess muss zusätzlich die entsprechenden Wortmarken auf der Perzeptionsseite auszeichnen, damit ein Monitoring-Prozess möglich wird. Man beachte, dass es Versprecher gibt, die ausdrucksseitig bedingt sind und zu einem Abbruch der Produktion mitten im Wort führen. Die Auszeichnung der Wortmarken auf der Perzeptionsseite muss also Phonem für Phonem erfolgen, was wiederum voraussetzt, dass es (in einem lokalistischen Repräsentationssystem) entsprechende neuronale Verbindungen gibt. Es muss

überlegt werden, was dieses Erfordernis für die These von der Trennung von Wortmarken-System und Sprechprogrammen bedeutet. Man beachte, dass die in Abbildung 4.2.4–1 eingezeichnete direkte Verbindung zwischen verbalem Eingangssystem und verbalem Ausgangssystem der Ansteuerung des Ausgangssystems von den Wortmarken aus dient (Engelkamp & Rummer, 1999: 160), von einem umgekehrten Vorgang ist nicht die Rede.

Ein zweites Problem entsteht dadurch, dass nicht ganz klar ist, wie das separate System der Sprechprogramme im Spracherwerbsprozess zustandekommt (Trial-and-Error für Wortformen und nicht nur für Laute?). Lernprobleme und Erklärungen für Lernprozesse sind auch hier wichtige Prüfsteine für die Qualität eines Modells sprachverarbeitender Prozesse.

Der Hinweis auf neuropsychologische Beobachtungen, die zur These der Trennung von Produktions- und Perzeptionssystem führen, ist wertvoll, solche Beobachtungen müssen aber in einen stimmigen Gesamtzusammenhang mit allen verfügbaren Informationen gebracht werden, auch solchen über mögliche Leistungen der neuronalen Architektur.

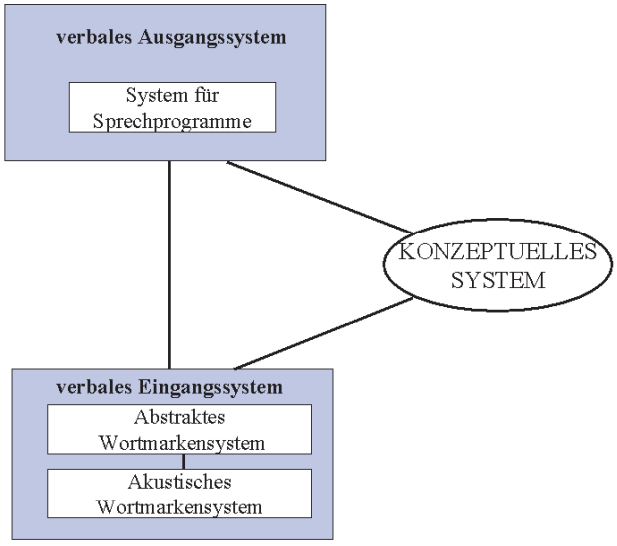

Abbildung 4.2.4–1: „Architektur des kognitiven Systems“ nach Engelkamp & Rummer (1999: 162), verkürzt auf den verbalen Bereich und ohne Lesen und Schreiben.

Es ist, unabhängig von der Frage der „apparativen“ Trennung von Perzeption und Produktion, sicherlich keine gute Forschungsstrategie, Modelle der Sprachproduktion unabhängig von Vorstellungen zum Funktionieren der Sprachperzeption zu entwickeln. Da die Sprachperzeption im Spracherwerb der Sprachproduktion vorausgeht, ist es eher möglich, sich mit der Perzeption in Isolation zu beschäftigen.

4.2.5 Repräsentation und Wirksamkeit „inhaltlicher“ Informationen

Die Frage nach der Natur lexikalischer Inhaltsseiten ist ein Stück weit beantwortet durch die Überlegungen zur Semantik in Teil 2, „Grundlagen“, Kapitel 2.4. Dort wird begründet, dass es nicht möglich ist, mit einer eigenständigen, ausschließlich sprachspezifischen Semantik zu rechnen. Das Hauptargument ist, dass die Verbindung einer sprachspezifischen Semantik mit Sinnesdaten nicht über Lernprozesse etabliert werden kann und Sinneswahrnehmung und Vorstellungsprozesse für sich genommen so strukturiert sind, wie es gängige Darstellungen der Semantik vorsehen, so dass die Annahme einer sprachspezifischen Semantik nur eine Doppelung von Informationen bedeuten würde. Ergänzend muss hinzugenommen werden, dass über den Spracherwerbsprozess eine Anpassung von Wahrnehmungskategorien (Bedeutungen) an die Konventionen der Sprachgemeinschaft möglich ist, und dass semantische Priming-Effekte (z. B. schnellere Reaktionszeiten, wenn auf einen verbalen Stimulus ein inhaltlich verwandter folgt) nicht durch direkte Verbindungen zwischen inhaltlichen Konzepten zu erklären sind, weil die dafür erforderliche Übertragung von Intensitätsinformationen in den zuständigen Kortexbereichen nicht möglich ist (vgl. unten 4.2.6).

Wenn die interne Strukturierung der lexikalischen Inhaltsseiten damit erklärt ist, bleibt im Zusammenhang mit der Behandlung lexikalischer Strukturen nur die Art der Verbindung der inhaltlichen Konzepte mit den sprachlichen Ausdrucksseiten als Problem übrig. Mindestens über *eine* neuronale Verbindung, die über Lernprozesse zu etablieren ist, müssen Ausdrucksseite und Inhaltsseite zusammenhängen. Es muss aber angesichts dieser Minimalvorstellung gefragt werden, ob durch eine solche Konstruktion gewährleistet werden kann, dass die von Marslen-Wilson (siehe oben Abschnitt 4.1.2) betonten frühen Wortentscheidungen zustande kommen können.

Die in verschiedenen Modellen des Lexikonzugriffs, auch in den früheren Arbeiten von Marslen-Wilson, vorgesehenen interaktiven Prozesse, das heißt die Möglichkeit der Rückwirkung „höherer“ Verarbeitungsergebnisse auf den

Fortgang der Analyse auf „niedrigerer“ Ebene, die dann auch wieder entsprechende neuronale Verbindungen voraussetzen würde, wird heute kontrovers diskutiert (vgl. z. B. Norris 2000). Eine solche Rückwirkung ist prinzipiell zu trennen von Vorgängen in der Sprachproduktion. Sie würde die Konzeption von Lernprozessen im lexikalischen Bereich erheblich erschweren. Es muss also vom Standpunkt lokalistischer neuronaler Modelle aus geklärt werden, ob die Annahme von Rückwirkungen, die z. B. von den Inhaltsseiten der Wörter ausgehen könnten, erforderlich ist.

Interessant, was die erforderlichen Verbindungen zwischen lexikalischen Ausdrucksseiten und Inhaltsseiten angeht, ist die Repräsentation inhaltlicher Mehrdeutigkeit. Es sind verschiedene Fälle zu unterscheiden, die in Abbildung 4.2.5–1 zusammengestellt sind.

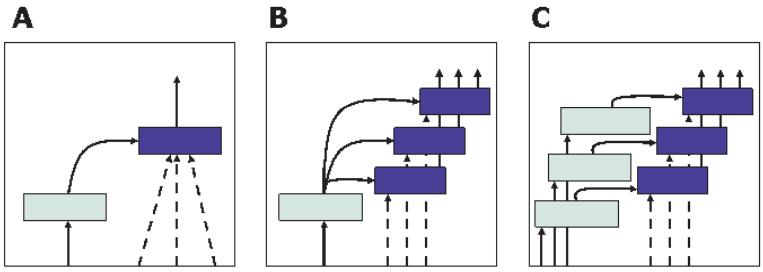

Abbildung 4.2.5–1: Drei Möglichkeiten der Repräsentation von „Mehrdeutigkeiten“. **A.** Mehrdeutigkeit durch Verknüpfung einer lexikalischen Ausdrucksseite mit einem „unscharfen“ inhaltlichen Konzept (Prototypizität). **B.** Mehrdeutigkeit durch Verknüpfung einer lexikalischen Ausdrucksseite mit mehreren inhaltlichen Konzepten. **C.** Mehrdeutigkeit durch Verknüpfung mehrerer *gleicher* lexikalischer Ausdrucksseiten mit je einem inhaltlichen Konzept.

Die Anordnung entspricht Abbildung 2.4.3–1 in Teil 2, „Grundlagen“, und stellt Verhältnisse dar, die zunächst die Verarbeitung im Rahmen der Sprachperzeption betreffen. Ein sprachlicher Input führt zu einem sukzessiven Abgleich der Ausdrucksseite, Details sind offen gelassen. Von der Repräsentation der Ausdrucksseite aus gibt es Verbindungen zu „inhaltlichen“ Repräsentationen, die über gestrichelt eingetragene Bahnen von den Sinnesorganen her definiert sind und entsprechende Kategorisierungen der Wahrnehmung leisten können. Damit ist es möglich, dass der Eindruck der Mehrdeutigkeit durch Prototypizität der inhaltlichen Kategorie entsteht, wie er in Teil 2 beschrieben und erklärt worden ist (**A**), nicht dadurch, dass die lexikalische Ausdrucksseite mit mehreren Kategorien verknüpft ist (**B** und **C**). Nur die

Strukturen **B** und **C** stellen Verbindungsmuster für die Bewältigung von sprachlicher Mehrdeutigkeit dar, die spezifisch für das sprachliche Lexikon sind und im Zusammenhang mit lexikalischen Strukturen als Möglichkeiten beachtet werden müssen. Beide setzen lokalistische Architekturprinzipien voraus. Architekturdetails könnten dazu führen, dass nur eine dieser beiden Möglichkeiten als realistisch akzeptiert werden kann, das ist aber nicht notwendig so.

4.2.6 Effekte in Experimenten

Es ist wichtig, festzuhalten, dass Lexikonmodelle zunächst die alltäglich und ohne die Lupe spezifischer Laborexperimente beobachtbaren Phänomene gewährleisten müssen. Dazu gehören z. B.:

- der unmittelbar und parallel zum Wahrnehmungsprozess erfolgende Verstehensprozess,
- die hohe zeitliche Effektivität,
- die Sicherheit des Produktionsprozesses,
- die Bewältigung akustischer Störungen,

und andere Leistungen, die wir als selbstverständlich und problemlos empfinden.

In den sehr zahlreichen Laboruntersuchungen, die Aussagen über die Struktur und Funktion des mentalen Lexikons liefern sollten, sind zusätzliche Effekte beobachtet worden, von denen die wichtigsten im Folgenden kurz charakterisiert werden sollen. Teilweise sind sie schon in den Beschreibungen der vorangegangenen Abschnitte angesprochen worden. Effekte, die zwar in wortbezogenen Experimenten beobachtet worden sind, die aber einigermaßen sicher postlexikalischen Prozessen zuzuordnen sind, und solche, die nur bei visuellen Stimuli auftreten, bleiben ausgeklammert.

Wortfrequenzeffekte

Wortfrequenzeffekte haben bei der Entwicklung psycholinguistischer Modelle des Lexikons eine wesentliche Rolle gespielt (siehe oben in Abschnitt 4.1.2). Häufigere Wörter führen in Worterkennungsexperimenten (sowohl bei visueller als auch bei akustischer Darbietung der Stimuli), aber auch beim Bildbenennen zu kürzeren Reaktionszeiten. Das wird nicht nur sichtbar, wenn man sehr häufige Wörter mit sehr seltenen vergleicht, sondern

auch abgestuft nach Häufigkeit. Es liegt auf der Hand, dass „Wortfrequenz“ zunächst individuell gesehen werden muss und dass das experimentelle Setting und die Interpretation von Ergebnissen dem Rechnung zu tragen hat.

Nachbarschaftseffekte

Es werden Effekte beobachtet, die davon abhängen, ob ein Wort mehr oder weniger ähnlich lautende Nachbarn hat. Sofort einzusehen ist, dass Wörter, die mehr Nachbarn haben, weniger genau identifiziert werden und dass die Häufigkeit der Nachbarn dabei eine Rolle spielt. Die Effekte auf Reaktionszeiten können, je nach experimentellem Design, in einer Beschleunigung oder einer Verlangsamung der Wahrnehmung oder des Benennens bestehen. Einen Überblick geben Lively, Pisoni & Goldinger (1994: 274 f.).

Priming-Effekte

Priming meint (im lexikalischen Bereich) allgemein die Erleichterung – was immer das heißen mag – der Verarbeitung eines Stimulus (target), der auf einen verwandten vorangehenden Stimulus (prime) folgt. Man kann drei Typen von Priming-Effekten unterscheiden:

- semantisches Priming, bei dem die Stimuli semantisch verwandt sind,
- phonologisches Priming, bei dem die Stimuli in ihrer Form verwandt sind, und
- „repetition priming“, das beobachtet wird, wenn ein Stimulus wiederholt wird.

Das Besondere beim „repetition priming“ ist, dass der Effekt auch auftritt, wenn ein größerer zeitlicher Abstand zwischen den Stimuli besteht. Es besteht offensichtlich eine Verwandtschaft mit den Wortfrequenzeffekten.

Die Standarderklärung für Priming-Effekte verwendet die Idee der „spreading activation“, also der Ausbreitung von Erregungsintensitäten in einem Netzwerk von Knoten, die den Stimuli oder deren Eigenschaften entsprechen. Diese Erklärung ist für Modelle, die den in Teil 2, „Grundlagen“ erarbeiteten Kriterien entsprechen, von vornherein schwierig, wie auch schon oben 4.2.5 festgestellt. Es muss nach anderen Erklärungsmöglichkeiten gesucht werden.

Phonemrestaurationseffekte

Ein klassisches Beispiel sind die Experimente, über die Warren (1970) berichtet: Wenn man in Wörtern einzelne Phoneme durch Geräusche ersetzt,

entsteht die „Illusion“, dass die ersetzten Phoneme doch vorhanden sind. Es gibt Beschränkungen für die Natur des Ersatzes, wie die unten beigefügten Tonbeispiele zeigen. Der Phonemrestaurationseffekt ist auch z. B. verantwortlich für die Unterscheidung von [f] und [s] bei schlechter Telefonqualität. In „interaktiven“ Modellen, in denen es nicht nur bottom-up gerichtete, sondern auch top-down gerichtete Einflüsse auf die Verarbeitung gibt, sind solche Effekte leichter zu erklären als in Modellen, in denen die Verarbeitungsebenen keine Rückwirkungsmöglichkeiten haben. Norris (2000) enthält einen Erklärungsversuch für diesen letzteren Fall.

Tonbeispiele:
Beispiel 1: *wissen,* unverändertes Original
Beispiel 2: *wissen,* [s] ist durch Husten ersetzt
Beispiel 3: *Wonne,* unverändertes Original
Beispiel 4: *Wonne,* [o] ist durch Husten ersetzt
Beispiel 5: *Wonne,* [o] ist durch [n] ersetzt
Beispiel 6: *Wonne,* [w] ist getilgt

Reizdegradierungseffekte

Reizdegradierung kann, beschränkt auf akustische Stimuli, Unterschiedliches meinen: Es können, wie bei den Experimenten zu den Phonemrestaurationseffekten, einzelne Phoneme durch Geräusche ersetzt oder verdeckt werden, es kann überhaupt Rauschen unterschiedlicher Qualität hinzugefügt oder die Lautstärke kann herabgesetzt werden. Schließlich können auch einfach unvollständige Stimuli verwendet werden. Reizdegradierung wird selbstverständlich zu einer Verschlechterung verschiedener Leistungen führen. Interessant ist es, die Auswirkungen in unterschiedlichen Fällen genauer zu analysieren. So hat in der Diskussion der Lexikonmodelle eine Rolle gespielt, dass eine Beeinträchtigung der Wortanfänge nicht zu einem generellen Scheitern des Lexikonabgleichs führen darf (siehe oben Abschnitt 4.1.2 zu TRACE).

Kontexteffekte

Kontexteffekte sind hier, in Abgrenzung zu den Primingeffekten, Einflüsse des Satz- bzw. Textkontexts oder des Situationskontexts auf die Worterkennung. Dass Kontexteinflüsse überhaupt wirksam sind, ist nicht die Frage, sondern es ist zu klären, ob man Kontexteffekte schon in den Vorgang des Abgleichs der lexikalischen Ausdrucksseiten einbeziehen kann, oder ob sie einem Prozess zuzuordnen sind, der erst danach, z. B. in einem Wettbewerb unter verschiedenen Wortkandidaten, wirksam wird. Charakteristisch für diese Problematik ist die Entwicklung des Kohortenmodells von Marslen-

Wilson (siehe oben 4.1.2), bei dem Kontexteinflüsse zunächst auf frühen Verarbeitungsstufen angenommen, und später letztlich dem postlexikalischen Bereich zugeordnet worden sind. Die Frage kann nicht als endgültig entschieden angesehen werden.

Es ist insgesamt eine beeindruckende Palette von Phänomenen durch ausgeklügelte Versuchsanordnungen beobachtet worden. Man muss sich bei der Bewertung allerdings bewusst halten, dass Feststellungen von gradierten Laboreffekten („mehr“, „besser“ usw.) statistische Aussagen sind, das heißt, sie beziehen sich auf Mittelwerte. Man kann nicht alle Variablen, die auf einen Messwert in einer konkreten Situation Einfluss haben könnten, kontrollieren. Wenn von einer kürzeren Reaktionszeit bei einer bestimmten experimentellen Aufgabenstellung die Rede ist, dann meint das einen entsprechenden Mittelwert über eine gewisse Anzahl von Versuchen gerechnet, nicht, dass jeder einzelne Versuch eine entsprechende Beobachtung erlaubt. Weiter ist zu beachten, dass, wie oben (4.2.1) schon angemerkt, Reaktionen eigene Produktionsvorgänge sind, und nicht einfach als Spiegel von Perzeptionsvorgängen interpretiert werden dürfen und auch nicht nur motorische Vorgänge im engeren Sinn einschließen.

Obwohl zunächst geklärt werden muss, wie die alltäglichen Beobachtungen auf der Basis von Annahmen über neuronal mögliche Prozesse zu verstehen sind, ist es doch auch erforderlich, wenigstens die experimentellen Standardeffekte, die vielfach repliziert worden sind, in die Argumentation einzubeziehen.

4.3 Lexikalische Ausdrucksseiten

Spracherwerbsprozesse, die zum Verstehen lexikalischer Einheiten führen, gehen der Produktion voraus. Strukturen, die Verstehen ermöglichen, sind offenbar unabhängig von der Produktion, in dem Sinne, dass Verstehen prinzipiell ohne Produktionsstrukuren funktionieren kann. Das ermöglicht es, die Aufklärung der Lexikonarchitektur zunächst im Wesentlichen unter Perzeptionsperspektive anzugehen. Verweise auf die Tradition der Sprachproduktionsmodelle haben zunächst nur die Funktion, die Argumentation zu stützen.

4.3.1 Grundprinzipien der Repräsentation von Phonemsequenzen

Nachdem in Abschnitt 4.2.2 festgestellt worden ist, dass Phoneme die sublexikalischen Einheiten sein müssen, die im Bereich des Lexikons als Input vorliegen, und dass das kürzeste phonologische Element im Verarbeitungsprozess durch einen einzelnen Impuls auf einer phonemspezifischen Verbindung ausgezeichnet ist, vereinfacht sich die Verarbeitung von lexikalischen Sequenzen wesentlich. Wenn verteilte Repräsentationen ausgeschlossen sind, können die mit solchen Repräsentationen arbeitenden Vorstellungen im Paradigma der künstlichen neuronalen Netze ausgeklammert werden. Die Idee, spezialisierte Zellen vorzusehen, die sowohl das Phonem als auch den Ort innerhalb einer Kette darstellen, in der Weise, wie das im Modell TRACE realisiert ist, scheidet schon aus einfachen Gründen der Plausibilität mehrerer Verarbeitungsdetails aus.

Alternative Vorschläge für den Bereich der Sequenzrepräsentation in der Sprachperzeption finden sich z. B. bei Pulvermüller (2002: Kapitel 6, 9 und 10). Einen relativ großen Raum nimmt die Diskussion von Vorschlägen ein, bei denen das Signal des Vorgängers in der Sequenz durch einen längeren Signalweg zeitlich verzögert wird, so dass Impulse des Vorgängers und

des Nachfolgers gleichzeitig in einer als „Sequenzdetektor“ wirkenden Einheit (Zelle) eintreffen, die dann durch einfache Summation das Vorliegen der Sequenz feststellen kann. Ein wesentlicher Nachteil ist, dass sich damit starre Zeitverhältnisse ergeben, sodass der variablen Geschwindigkeit des sprachlichen Inputs nicht Rechnung getragen werden kann. Ein weiteres Problem ist, dass doch wohl dergleichen Strukturen (z. B. unterschiedliche Leitungslängen oder „Delay-Neuronen“) nicht über Lernvorgänge etabliert werden können, sondern nur in angeborenen Teilnetzen möglich sind. Niemand zweifelt aber daran, dass das Lexikon der sprachlichen Ausdrucksseiten durch Lernvorgänge zustande kommt.

Während also solche, mit zeitlicher Verzögerung arbeitenden Modelle – und einige weitere, bei Pulvermüller (2002, Kapitel 6.1) kurz angesprochene, aber einen Simulationsversuch mit biologischen Neuronen nicht überstehende Konstruktionen – aus grundsätzlichen Gründen abzulehnen sind, ist ein anderer, bei Pulvermüller (2002: 177 f.) referierter Vorschlag wenigstens prinzipiell auch mit den in Teil 2, „Grundlagen“, erarbeiteten Bedingungen für neuronale Repräsentationen und Prozesse vereinbar (die allerdings nicht den von Pulvermüller vorausgesetzten Strukturen entsprechen). Es wird angenommen, dass die Sequenzelemente nicht verschieden verzögerte, sondern verschieden starke Verbindungen mit einem „Sequenzdetektor“ eingehen, das heißt für uns: mit einer Großmutterzelle, die die Sequenz repräsentiert. Die Verbindungsstärken sind so geordnet, dass das erste Sequenzelement die schwächste, das letzte Sequenzelement die stärkste Verbindung hat. Verbindungsstärken können prinzipiell durch Lernvorgänge verändert werden, es ist allerdings nicht bekannt, wie Lernvorgänge für diesen speziellen Fall aussehen könnten. Abbildung 4.3.1–1 zeigt unter **A** eine entsprechende, sehr einfache Anordnung.

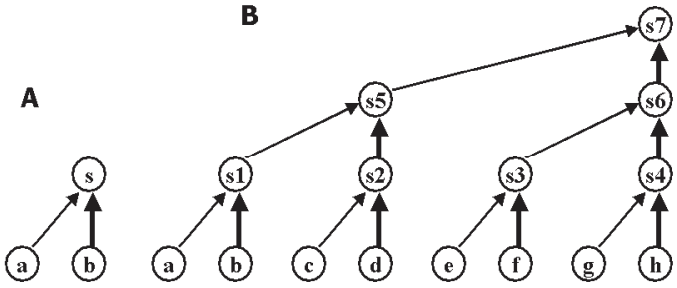

Abbildung 4.3.1–1: Architekturen mit Sequenzdetektoren. Die erforderliche Abstufung der Synapsengewichte ist durch die Pfeilstärke ausgedrückt. **A.** Der einfachste Fall. **B.** Eine ausgewogene Hierarchie von Sequenzdetektoren.

Simulationen zeigen sofort auch die Schwächen dieser Konstruktion: Es werden zwar die Sequenzen a-b und b-a unterschieden, nur die Abfolge a-b führt zum Feuern des Sequenzdetektors *s* – allerdings unter der Voraussetzung, dass die Abstände der Sequenzelemente entsprechend groß sind; der Unterschied wird verwischt, wenn b ausreichend kurz vor a erscheint. Aber *s* feuert auch bei der Sequenz b-b und es muss darauf geachtet werden, dass zwischen den einzelnen Sequenzen ausreichend Zeit vergeht, damit die Potenziale, die im Sequenzdetektor ggf. stehen bleiben, wenn nicht die Sequenz vorliegt, auf die er spezialisiert ist, ausreichend weit abklingen.

Simulationen:
Sequenz a-b mit ausreichendem zeitlichem Abstand
Sequenz b-a mit ausreichendem zeitlichem Abstand
Sequenz b-b
Sequenz b-a mit zu kurzem zeitlichem Abstand
Sequenz a-b-c mit Permutationen.

Die auf dem Bildschirm verwendete Symbolik wird in Teil 2, „Grundlagen“, Abbildung 2.1.4–2, erläutert.

Es ist zwar möglich, mehr als zwei Eingänge in einem Sequenzdetektor zu verknüpfen, es muss dann aber die Weiterverarbeitung warten, bis eine solche Sequenz vollständig eingegeben ist, was den Befunden widerspricht, die zur Kohortentheorie geführt haben. Also wird man Hierarchien von Sequenzdetektoren annehmen. Ausgewogene (symmetrische) Hierarchien nach Abbildung 4.3.1–1 **B** führen zu dem Problem, dass die Teilsequenzen in den meisten Fällen keine brauchbaren Hypothesen für die Weiterverarbeitung darstellen und es zusätzlich erforderlich ist, anzunehmen, dass die Zeitdauer für die Wirksamkeit der EPSPs in den Sequenzdetektoren unterschiedlich ist und relativ groß werden kann.

Die sinnvollste Lösung im Hinblick auf die laufende Weiterleitung von Teilergebnissen während des Verarbeitungsprozesses ist die der Abbildung 4.3.1–2, zusätzlich entstehen hier nicht die angesprochenen Zeitprobleme. Jeder Sequenzdetektor kann als Repräsentant einer Teilsequenz betrachtet werden, die vom Sequenzanfang bis zu dem gerade bearbeiteten und erkannten Sequenzelement reicht. Allerdings bleiben die schon oben für die elementarste Anordnung festgestellten Mehrdeutigkeiten. Bei der Struktur der Abbildung 4.3.2–1 wären die Sequenzen a-b-c-d-e-f-g-h und b-b-d-d-f-f-g-h und ähnliche Varianten ununterscheidbar. Eine Repräsentation dieser Art für die Sequenz a-b-c-a-b-d würde endlose Wiederholungen der Teilkette a-b-c zulassen so z. B. die Sequenz a-b-c-a-b-c-a-b-d. Dergleichen ist für ein Modell der Verarbeitung lexikalischer Sequenzen im Rahmen des Sprachverstehens und selbst dann, wenn man gewisse Mehrdeutigkeiten zugesteht, nicht tolerabel.

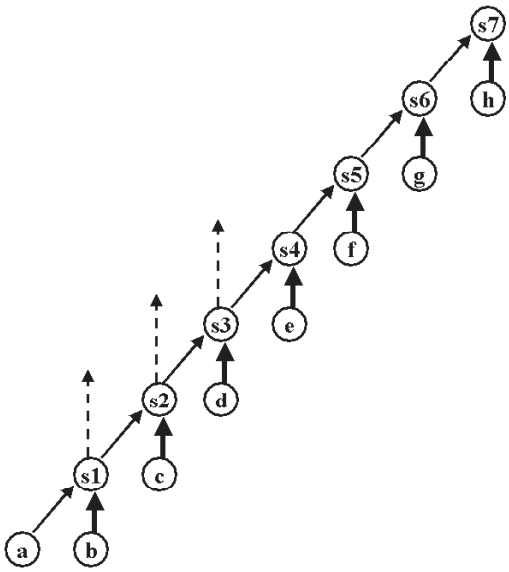

Abbildung 4.3.1–2: Hierarchie von Sequenzdetektoren mit der Möglichkeit der Reaktion auf sinnvolle (der Kohortentheorie entsprechende) Teilsequenzen. Die von *s1* ausgehende gestrichelte Verbindung repräsentiert die Teilkette a-b, die von *s2* ausgehende Verbindung die Teilkette a-b-c usw., nicht alle gestrichelten Verbindungen sind eingezeichnet.

Simulation:
Endlos wiederholbare Teilketten in einem Netz für
die Kette a-b-c-a-b-d.
Die sich wiederholenden Segmente a und b sind der Übersichtlichkeit
halber auf jeweils zwei Zellen a1 und a2 bzw. b1 und b2 verteilt,
die gleichzeitig extern erregt werden. Ab Zeittakt 1: a-b-c-a-b-c-a-b-d,
ab Zeittakt 600: b-b-d-b-b-d-b-b-d.

Wenn man genauer überlegt, wie die Probleme dieser Architekturvorstellung zustande kommen, stößt man vor allem auf die störende Wirkung des „Stehenbleibens“ unerwünschter Potenziale, vor allem von Potenzialen, die eigentlich nur ein schwaches EPSP, falls vorhanden, zu einem überschwelligen ergänzen sollten. Das erinnert an die Bedingungen, die in Teil 2, „Grundlagen“, Kapitel 2.4.6, im Zusammenhang mit der Darstellung von Kontextphänomenen diskutiert worden sind. Von dieser Spur ausgehend kann man sofort erkennen, dass lexikalische Sequenzen tatsächlich ein Sonderfall von Kontextsequenzen sind: Auf die Wahrnehmung einer lexikalischen Teil-

kette folgt die Erwartung einer bestimmten Fortsetzung oder auch mehrerer Fortsetzungen, bzw. die Erwartung des Kettenendes.

In der Architektur von Abbildung 4.3.1–2 hat der Sequenzanfang insofern eine Sonderstellung, als er nicht „erwartet“ wird, sondern einfach durch das Vorkommen des entsprechenden Sequenzelements aktiviert wird. Das führt dazu, dass, in einem Lexikon, in dem fast alle phonologischen Einheiten einer Sprache im Wortanlaut vorkommen, nahezu jeder Input eines Lauts im Verstehensprozess einen neuen Lexikoneintrag startet. Dieses Verhalten ist in Lexikonmodellen nicht unbekannt, man vergleiche z. B Norris (1994). Es ist aber doch so, dass, um ein beliebiges Beispiel zu geben, bei einem eindeutig erkannten Sequenzstart mit [ta] bei Eingabe von [k] (für das Wort *Takt*) nicht auch *Karl, Kunst, Kegel, Kompost* usw. aktiviert werden sollten (wieder im Hinblick auf Kohorteneffekte). Erwartungen eines Sequenzstarts sind nicht sinnvoll, solange die bisher erkannte Kette auch bei Berücksichtigung aller Mehrdeutigkeiten keine (morphologisch) vollständige Sequenz bildet (siehe auch oben Abschnitt 4.1.1 zu Klatt). Wenn also auch der Sequenzstart und damit die Gesamtsequenz als erwartet zu gelten hat, ergibt sich, aufgrund der Argumente, die in Teil 2, Kapitel 2.4.6 entwickelt worden sind, für die in Abbildung 4.3.1–2 dargestellte Sequenz die in Abbildung 4.3.1–3 wiedergegebene Architektur.

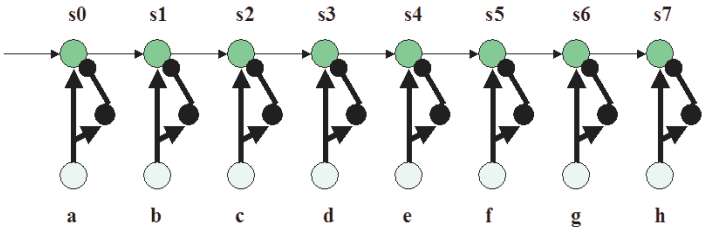

Abbildung 4.3.1–3: Lexikalische Ausdruckssequenz als Kontextsequenz, Verbindungsstärke ist durch Pfeildicke dargestellt.

Eine einzelne Einheit innerhalb der Sequenz wird durch drei Zellen realisiert, die jeweils unterschiedliche Eigenschaften haben müssen:

- Eine Zelle, die dem Sequenzdetektor entspricht und in Kapitel 2.4.6 als „sequenzenbildende Zelle“ bezeichnet worden ist. Sie besitzt durch Lernvorgänge verstärkbare Verbindungen zu anderen sequenzenbildenden Zellen. Das EPSP der sequenzenbildenden Zellen hat eine Abnahmerate, die so gering ist, dass Effekte aufeinanderfolgender Eingabephoneme verknüpft werden können, was einer Zeitdauer von größenordnungsmäßig 100 ms entspricht.

- Eine Zelle, die Inputs von prälexikalischen Strukturen erhält, und die als „instanzenbildende Zelle“ bezeichnet worden ist. Da diese Zelle keine Sequenzen abbilden soll, muss ihr EPSP eine entsprechend kurze Dauer haben, vielleicht weniger als 30 ms.

- Eine hemmende Zelle, deren hemmende Wirkung auf die sequenzenbildende Zelle ausreichend groß sein muss, die aber selbst sonst keine Berechnungen durchführt. Der hemmende Impuls kommt an der sequenzenbildenden Zelle schon wegen des Umwegs etwas später an als der erregende Impuls von der instanzenbildenden Zelle aus.

Die Zelldifferenzierung wird in Kapitel 2.4.6 zum Teil auch durch Hinweis auf Erfordernisse von Lernprozessen begründet. Für die Verschaltung gilt, dass die Verbindungen, die von der instanzenbildenden Zelle ausgehen, nicht durch Lernvorgänge aufgebaut werden. Es handelt sich um eine überschwellige Verbindung mit der hemmenden Zelle, und eine sehr starke, aber unterschwellige Verbindung mit der sequenzenbildenden Zelle. Die hemmende Verbindung ist ebenfalls nicht als Produkt von Lernprozessen zu erklären.

Der Unterschied im zeitlichen Verhalten von sequenzenbildender und instanzenbildender Zelle ist gleichbedeutend mit dem Unterschied von Gleichzeitigkeit und Abfolge. Ereignisse, die weniger als 30 ms auseinander liegen, gelten, wenn man diese Grenze annimmt, als gleichzeitig. Unter derselben Annahme wird dann vorausgesetzt, dass in einem Bereich mit einem gewissen zeitlichen Abstand, also z. B. größer 40 ms und kleiner 90 ms, jeder phonologische Bestandteil einer Inputsequenz in das Lexikon durch ein einzelnes Aktionspotenzial repräsentiert ist. Nachdem Langvokale im Deutschen typischerweise länger als 90 ms dauern und, wie Konsonanten mit Ausnahme von Verschlusslauten auch, beliebig längbar sind, muss für Langvokale und die anderen Längen gelten, dass sie mehr als ein Aktionspotenzial auslösen, für Langvokale sind im Unterschied zu Kurzvokalen mindestens zwei Aktionspotenziale erforderlich.

Die Zeitverhältnisse in einem Einzelfall werden in dem Spektrogramm der Abbildung 4.3.1–4 deutlich, in dem das Wort *Banane* in sehr schneller Sprechweise (vgl. das entsprechende Tonbeispiel) dargestellt ist. Jeder in das Spektrogramm eingetragene rote Punkt entspricht einem Aktionspotenzial, das aufgrund eines identifizierten phonologischen Segments abgegeben wird. Wenn man die Zeitspanne zwischen zwei Aktionspotenzialen gleichmäßig auf ca. 45 ms einstellt, trifft man während des Inputs immer auf einen Punkt, der tatsächlich in ein phonologisches Segment fällt. Das kurze /a/ produziert nur ein einziges Aktionspotenzial, das lange /a:/ zwei Aktionspotenziale. Die Zeitspanne von 45 ms, wenn man die sehr schnelle

Sprechgeschwindigkeit dieses Beispiels schon als einen Grenzfall betrachtet, wird durch den ersten Kurzvokal festgelegt.

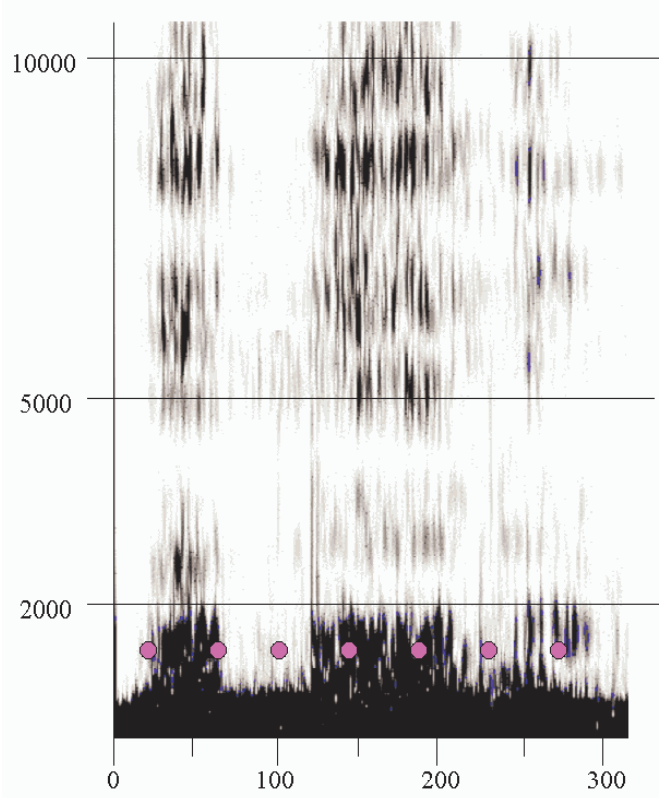

Abbildung 4.3.1–4: Spektrogramm von *Banane* in sehr schneller Sprechweise.

Tonbeispiel:
Banane wie im Spektrogramm von Abbildung 4.3.1–4 dargestellt

Die Konsequenzen für die verarbeitende Architektur müssen sein: Um einen Langvokal sozusagen zu „erzwingen“, müssen zwei Sequenzelemente (also zwei entsprechende Sätze von Neuronen) aufeinanderfolgen. Um eine beliebige Längung zu ermöglichen, muss mindestens eines dieser Sequenzelemente, nach der Argumentation in Teil 2, Kapitel 2.4.6, eine Verbindung auf sich selbst zurück haben, die ein unterschwelliges(!) Erwartungspotenzial erzeugt. Längbare Konsonanten können durch ein einzelnes Sequenzelement

mit einer Verbindung auf sich selbst zurück dargestellt werden. Kurzvokalrepräsentationen, die die Kürze erzwingen sollen, dürfen keine solche Verbindung haben. Das ergibt für das Wort *Banane* die Sequenzarchitektur von Abbildung 4.3.1–5.

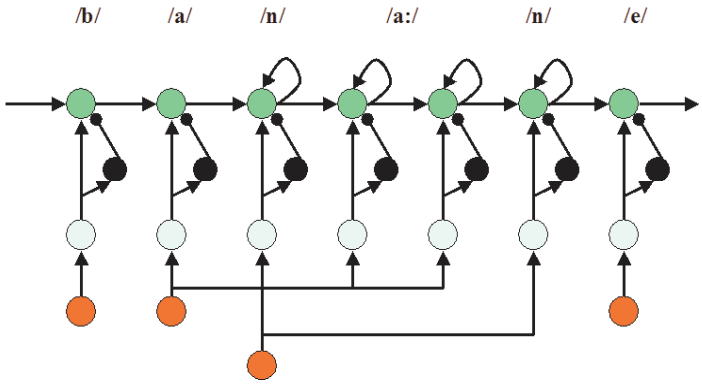

Abbildung 4.3.1–5: Lexikalische Ausdruckssequenz für das Wort /bana:ne/.

Auf diese Lösung des Problems der Sequenzrepräsentation ist seit Kochendörfer (1994) immer wieder hingewiesen worden. Mehrdeutigkeiten, die zu Störungen des Verarbeitungsprozesses führen würden, entstehen hier nicht mehr. Die Verdoppelung einer Eingabe z. B. hat banalerweise deshalb keine Störung zur Folge, weil ein dadurch bewirkter unterschwelliger Effekt in einer sequenzenbildenden Zelle über die Hemmung sofort wieder gelöscht wird, und auch die Wiederholung des Starts durch Eingabe des ersten Sequenzelements ist wegen des fehlenden Erwartungspotenzials nicht möglich.

Eine wichtige Eigenschaft dieser Konstruktion ist, dass die Elemente in den Ausdruckssequenzen „Tokens“ von Lauten sind. Die beiden /n/ und die /a/-Phoneme im Beispiel /bana:ne/ sind Tokens eines und desselben „Types“. In einigen Modellen der lexikalischen Produktion (z. B. Dell, 1986; Schade, 1992) wird, motiviert durch den Versuch, Versprecherdaten zu erklären, angenommen, dass die Festlegung der zu äußernden phonologischen Einheiten und die Festlegung ihrer Reihenfolge verschiedene, auch apparativ trennbare Prozesse darstellen. Bei Dell (1986) werden selektierte Einheiten strukturellen „slots“ zugewiesen, die die Position in der Äußerung bestimmen. Mehrfaches Vorkommen eines Lauts innerhalb eines Worts erfordert nicht, dass mehrere Exemplare (Tokens) dieses Lauts vorhanden sind (Dell, 1986, 288). Bei Schade (1992) entspricht dem die Unterscheidung von „Normknoten“ und „Kontrollknoten“. Die „wörtliche“ Umsetzung dieser Konstruktion für den Vorgang der Sprachperzeption führt zu Problemen

bei der Unterscheidung von Wortformen, die in Kochendörfer (1997: 260 ff.) etwas ausführlicher diskutiert sind. Die Möglichkeit von Lernvorgängen, die zu den erforderlichen Strukturen führen würden, ist ungeklärt. Es müssten Großmutterzellen definiert werden, die nicht *Elementen* von Ausdrucksseiten, sondern *Positionen* von Elementen entsprechen! (Andere Möglichkeiten, die Zahl der erforderlichen Tokens zu reduzieren, werden im folgenden Abschnitt besprochen.)

Selbst wenn der Input in eine der instanzenbildenden Zellen in einem Burst von Impulsen (der eine entsprechend hohe Frequenz haben müsste) bestehen würde, würde die entsprechende sequenzenbildende Zelle maximal einen einzelnen Impuls abgeben, es sei denn, das Erwartungs-EPSP dieser sequenzenbildenden Zelle würde mit entsprechender Frequenz erneuert. Versucht man, diese zusätzliche Bedingung herzustellen, muss man darauf achten, dass diese Auffrischung tatsächlich noch die Funktion der „Erwartung“ behält und nicht für sich allein ein Aktionspotenzial auf der sequenzenbildenden Zelle auslöst, auf die sie einwirkt. Wenn man gewährleisten möchte, dass ein Burst in der gleichen Länge wie das Inputsignal von dieser Zelle abgegeben wird, müssen die „horizontalen“ und die „vertikalen“ Inputs dieser Zelle synchronisiert gedacht werden. Diese Annahme kann schon allein deshalb nicht realistisch sein, weil die Produktion des Bursts im Sequenzvorgänger der Aktivität des Frequenznachfolgers definitionsgemäß vorausgehen muss. Die Konsequenz aus dieser Überlegung ist, dass die Verarbeitung einzelner Impulse, deren Erzeugung von dem phonologischen Input gesteuert wird, die Basis für die Sequenzerkennung ist. Man beachte, dass dieses Ergebnis den in Teil 2, „Grundlagen“, für neuronale Lernvorgänge festgestellten Voraussetzungen entspricht.

4.3.2 Verbände von Ausdrucksseiten

Im vorigen Abschnitt ist die Darstellung einzelner lexikalischer Ausdrucksseiten diskutiert worden, und es ist offen gelassen worden, ob es auf der ausdrucksseitigen Ebene, also nicht nur ggf. inhaltsseitig, Beziehungen zwischen den Lexikoneinträgen gibt. Die Klärung dieser Frage ist Gegenstand der folgenden Überlegungen, die dann allerdings erst in Abschnitt 4.3.3 zu einem Abschluß kommen.

Die gedanklich einfachste Vorstellung von einem Lexikon der sprachlichen Ausdrucksseiten ist wahrscheinlich die der Liste. Jeder Lexikoneintrag wird als Sequenz verstanden, die unabhängig als Element der Liste erscheint. Für die Wörter /ban/ (*Bann*), /ba:n/ (*Bahn*), /bana:ne/ (*Banane*), /an/ (*an*), /anne/ (*Anne*), /a:ne/ (*Ahne*) und /na:be/ (*Nabe*) wäre also die folgende

Liste von Sequenzen vorauszusetzen:

b - a - n
b - a - a - n
b - a - n - a - a - n - e
a - n
a - n - e
a - a - n - e
n - a - a - b - e

Die Simulation zeigt, dass dieses kleine (Teil-)Lexikon problemlos funktioniert. Bei mehrdeutigen Wortanfängen werden einfach mehrere Ketten gleichzeitig aktiviert und die Menge der Kandidaten wird, gemäß den Beobachtungen, die zur Kohortentheorie geführt haben, mit Fortschreiten des Inputs immer mehr reduziert.

Simulation:
Listenförmige Repräsentation lexikalischer Ausdrucksseiten. Es werden nacheinander die oben genannten Wörter eingegeben.

Diese Repräsentationsform verwendet relativ viele Tokens, oder etwas allgemeiner, unter Berücksichtigung der Möglichkeit von Modifikationen einzelner Tokens gegenüber dem zugrundeliegenden Type, „Instanzen“ von Phonemen. Eine Möglichkeit, zu einer Verminderung dieses Aufwands zu kommen, ist es, „Verklebungen“ im Sinne von Schukat-Talamazzini (siehe oben Abschnitt 4.1.1) vorzusehen. Ein mögliches Netz mit einer Maximalzahl von Verklebungen ist in Abbildung 4.3.2–1 gezeigt.

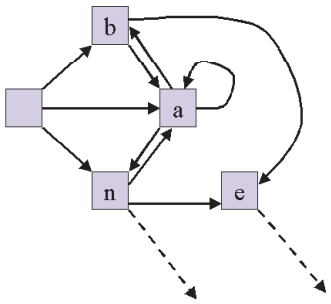

Abbildung 4.3.2–1: Lexikon mit maximaler Zahl von Verklebungen. Die gestrichelten Pfeile markieren mögliche Kettenenden.

Auch dieses Netz funktioniert prinzipiell – allerdings mit der Konsequenz, dass außer den eigentlich darzustellenden Sequenzen auch zusätzliche, unerwünschte Sequenzen zugelassen werden, was für ein natürlichsprachliches Lexikon in dem zu erwartenden Umfang nicht vertretbar ist.

Simulation:
Maximal „verklebte“ Repräsentation lexikalischer Ausdrucksseiten. Es werden wieder nacheinander die oben genannten Wörter eingegeben, anschließend die „Nichtwörter“ /ne/, /bane/ und /ena/, von denen die ersten beiden als korrekt akzeptiert werden (zu einem gültigen Kettenende führen), aber nicht das dritte. Die Sequenzverbindungen werden auf dem Bildschirm teilweise verdeckt, entsprechen aber für die Simulation der Abbildung 4.3.2–1.

Es muß also mindestens dafür Sorge getragen werden, dass nur die vorgesehenen Eingabewörter erfolgreich abgearbeitet werden können. Das ist in dem Netz der Abbildung 4.3.2–2 geschehen.

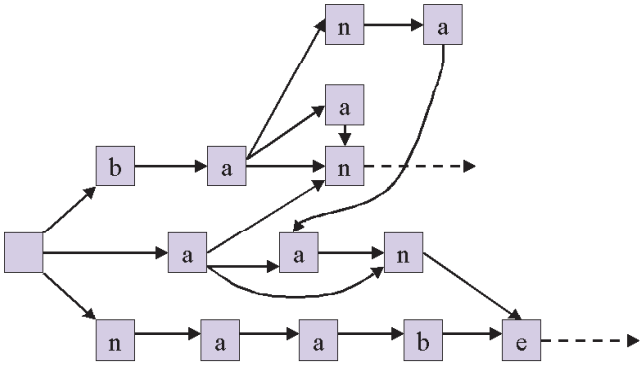

Abbildung 4.3.2–2: Lexikon mit Verklebungen, das nur die gewünschten Einträge gewährleistet. Die Kettenenden sind, wie in der vorigen Abbildung, durch gestrichelte Pfeile markiert.

Simulation:
Befriedigend funktionierende „verklebte“ Repräsentation lexikalischer Ausdrucksseiten. Der Input entspricht dem der vorigen Simulation. Die Verbindung *sa4* → *sn3* ist verdeckt.

Eine weitere Überlegung betrifft jetzt die Funktion dieses Netzes im Hinblick auf die daran anschließenden Verarbeitungsvorgänge im Sprachverstehensprozess. Der Sinn des Lexikonabgleichs besteht ja darin, dass eine (eindeutige) Eingabekette im Idealfall eindeutig erkannt wird und ausschließlich

zur Zuordnung einer ihr entsprechenden inhaltlichen Repräsentation führt. Bei der Konstruktion der Abbildung 4.3.2–2 sind aber gerade die Pfadenden mehrdeutig. Es ist ganz offenbar nicht sinnvoll, anzunehmen, dass prinzipiell Pfade, die zunächst getrennt sind, wieder zusammenführen können. Eine entsprechende Korrektur ergibt dann das Netz der Abbildung 4.3.2–3.

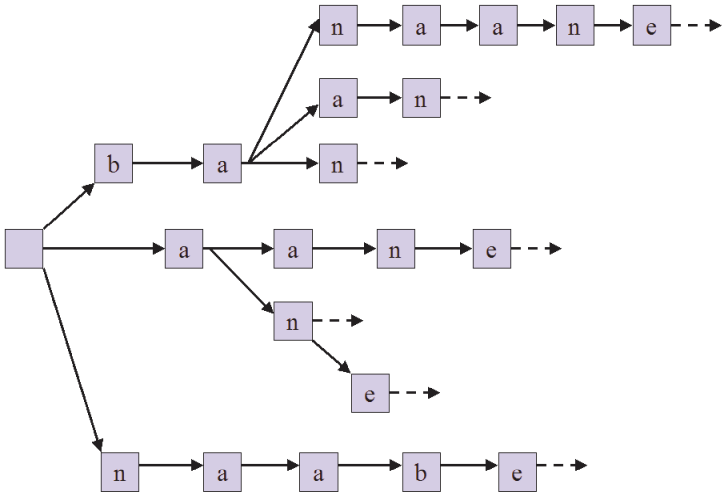

Abbildung 4.3.2–3: Lexikon mit Verklebungen der Kettenanfänge und eindeutigen Kettenenden.

Simulation:
Repräsentation mit verklebten Kettenanfängen, Kettenenden sind eindeutig. Der Input entspricht dem der vorigen Simulation.

Diese Konstruktion entspricht in ihrer Funktion unter allen bisher diskutierten Bedingungen den an den natürlichsprachlichen Lexikonabgleich zu stellenden Erwartungen. Das gilt auch für die Toleranz gegenüber der zeitlichen Variabilität des Inputs. In den folgenden Simulationen ist der Abstand zwischen zwei phonologischen Inputimpulsen einmal auf 50 Zeittakte (sehr rasch) und einmal auf 80 Zeittakte festgelegt, alle Netzparameter werden beibehalten.

Simulationen:
Zeitliche Variabilität des Inputs (Beispiel: *Banane*):
1. Sehr rasch.
2. Sehr langsam.

Da die Gesamtstruktur verwandt ist mit einer Trie-Struktur wie in 4.1.1 beschrieben, kann man die Segmentierung einer unsegmentierten lautlichen Eingabekette in lexikalische Bestandteile dadurch erreichen, dass die Pfadenden mit der Zelle verknüpft werden, die den Startkontext für das Lexikon erzeugt. Das entspricht bis zu einem gewissen Grad auch dem ebenfalls in 4.1.1 besprochenen Lexikon bei Klatt. Die besondere Funktion des Lexikonstarts ist oben in 4.3.1 begründet worden. In der folgenden Simulation ist das Netz der Abbildung 4.3.2–3 durch Rückverbindungen der Kettenenden mit der „Startzelle“ *st* ergänzt. Der Input /bana:ne/ führt zur Identifikation der Ketten /bana:ne/, /ban/ (*Bann*) und /a:ne/ (*Ahne*). Wenn das Lexikon, wie realistischerweise anzunehmen, mit einer Syntaxkomponente verknüpft wird, müssen die Rückverbindungen durch die Syntax vermittelt sein.

Simulation:
Lexikonstruktur mit Segmentierungsfunktion.
Input ist /bana:ne/.

Die Lexikonstruktur, wenn man den Zellaufwand betrachtet, zeigt eine gewisse Komplexität, die unvermeidbar ist, und die vor allem dadurch entsteht, dass nicht nur eine Datenmenge zu repräsentieren ist, sondern, wie für ein parallelverarbeitendes System notwendig, die zur Verarbeitung erforderlichen Prozesse gewährleistet werden müssen (keine Trennung von Speicherung und Verarbeitung). Im lexikalischen Bereich kann besonders anschaulich die unverzichtbare Funktion der hemmenden Zellen in der Kontextverarbeitung demonstriert werden. Wenn man die Wirkung der Hemmung reduziert oder ganz eliminiert, entstehen unerwünschte Mehrdeutigkeitseffekte. Es sind Effekte, die auch in anderen Zusammenhängen die Kontextverarbeitung stören würden.

Simulation:
Lexikonstruktur ohne Hemmung. (Die hemmenden Zellen sind zwar vorhanden, aber nicht funktionsfähig. Der Input entspricht der oben verwendeten Wortliste /ban/, /ba:n/ usw., einschließlich der Nichtwörter.)

4.3.3 Lernprozesse

Wenn Lernprozesse in konnektionistischen Modellen mit strikt verteilten Repräsentationen realisiert werden, tritt schon allein definitionsgemäß ein Phänomen auf, das als „catastrophic interference“ bezeichnet wird. Gemeint ist, dass eine bereits gelernte Zuordnung von Input-Repräsentation und Output-Repräsentation durch das Lernen der jeweils folgenden Zuordnung

in mehr oder weniger großem Umfang gestört wird. Diese Eigenschaft führt dazu, dass in der konnektionistischen Tradition standardmäßig Lernprozeduren verwendet werden, bei denen eine komplette Liste von Zuordnungen mehr oder weniger oft durchlaufen wird, bis für alle Zuordnungen eine als ausreichend angesehene Wiedergabeleistung erreicht ist (Abbildung 4.3.3–1 **A**). Eine solche Lernstrategie ist für den Aufbau des natürlichsprachlichen Lexikons nicht realistisch. Lexikalisches Lernen geschieht lebenslang, am ehesten wohl nach dem Muster von Abbildung 4.3.3–1 **C**. Das Hinzufügen eines neuen Lexikoneintrags darf jedenfalls nicht zur Folge haben, dass der bisherige Bestand ernsthaft in Frage gestellt wird.

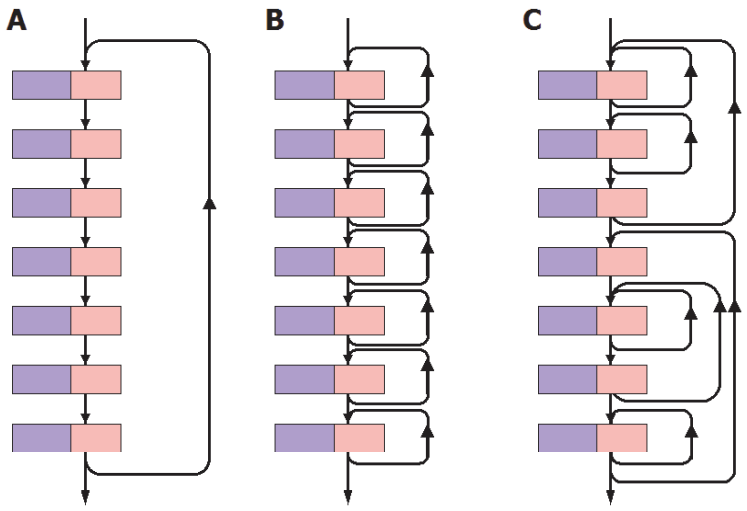

Abbildung 4.3.3–1: Typen von Lernzyklen. **A**: Typische Lernzyklen in konnektionistischen Modellen. **B**: Lernzyklen, bei denen Wort für Wort gelernt wird. **C**: Beispiel einer komplexen Verschachtelung von Lernzyklen, die vielleicht am ehesten der biologischen Realität entspricht.

Wie für alle biologisch realistischen Lernprozesse im Kortex gilt auch für lexikalisches Lernen, dass es „unüberwacht“ erfolgen muss. Das bedeutet, dass alle Komponenten und Voraussetzungen des Lernvorgangs bei den einzelnen Zellen lokalisiert sind, die die zu lernende Struktur bilden. Sowohl die Vermeidung von catastrophic interference als auch die Notwendigkeit unüberwachter Lernprozesse (man vgl. dazu die Auseinandersetzung in Teil 2, „Grundlagen“ Kapitel 2.3.2) sind wesentliche Argumente, die gegen die Annahme verteilter Repräsentationen und für die Richtigkeit lokalistischer Architekturen sprechen.

Nicht alle Bestandteile der im vorigen Abschnitt konstruierten Lexikonstrukturen können bzw. müssen durch Lernprozesse entstehen; das ist oben in Abschnitt 4.3.1 mit Bezug auf Teil 2, Abschnitt 2.4.6 schon angedeutet worden. Die Verbindungen innerhalb eines Sequenzelements sind bei allen Sequenzelementen gleichartig. Ein Sequenzelement wird gebildet durch eine instanzenbildende Zelle, eine hemmende Zelle und eine sequenzenbildende Zelle. Die instanzenbildende Zelle hat eine sehr starke erregende Verbindung auf die sequenzenbildende Zelle. Die Synapseneffektivität bleibt um einen geringen Betrag (in den Simulationen 2%) unter dem Wert, bei dem eine Erregung ein Aktionspotenzial in der sequenzenbildenden Zelle auslösen würde. Eine zweite Verbindung der instanzenbildenden Zelle hat überschwellige erregende Wirkung auf ein hemmendes Interneuron, das seinerseits EPSPs auf der zugeordneten sequenzenbildenden Zelle löscht. Es ist schwer vorstellbar, wie diese hemmende Verbindung durch einen unüberwachten Lernprozess aufgebaut werden könnte. Die gesamte Anordnung trägt nicht zur Variabilität der unterschiedlichen einzelsprachlichen Lexika bei und kann für angeboren gehalten werden.

Über diese fest verschalteten Grundelemente hinaus werden alle weiteren Strukturen durch Lernprozesse aufgebaut, die im Verstärken vorhandener Verbindungen bestehen. Der Lernvorgang setzt also ein Überangebot zunächst schwacher, verstärkbarer Verbindungen voraus. Es müssen solche Verbindungen vorhanden sein zwischen phonologischen „Types“, die die Ausgabe der Phonologie realisieren, und den instanzenbildenden Zellen der lexikalischen Sequenzelemente. Eine zweite Klasse von potenziellen Verbindungen wird zwischen den sequenzenbildenden Zellen vorausgesetzt. Es kann also nicht so sein, dass vor einem Lernprozess beliebige Zellen mit beliebigen Zellen schwach verknüpft sind, sondern Verbindungsmuster sind prinzipiell spezifisch für einzelne Zelltypen.

Es sind in Teil 2, „Grundlagen“, zwei Lernprozesse unterschieden worden, die zu unterschiedlichen Verknüpfungstypen führen, die als UND-Verknüpfung einerseits bzw. ODER-Verknüpfung andererseits aufgefaßt werden können. UND-Verknüpfungen entstehen dadurch, dass eine Zelle zunächst nicht einer bestimmten Inputkombination zugeordnet ist und dann, nach Verstärken entsprechender Synapsen, aufgrund eines Inputs feuert. Das Feuern beendet den Lernprozess an der Zelle und führt außerdem zur Fixierung der beteiligten Synapsen, das heißt zu einer Verlangsamung des Abbaus der erreichten Synapsenstärken. Diese Vorgänge sind genauer in Teil 2, „Grundlagen“, Abschnitt 2.3.2 beschrieben und begründet. Sie sind die Grundlage jeder Konzeptbildung. Eine ODER-Verknüpfung entsteht dadurch, dass eine Zelle über eine feste Verbindung zum Feuern gebracht und damit über

eine kurze Zeitspanne „lernbereit“ gemacht wird. Potenzielle Verbindungen, die in dieser Zeitspanne erregt werden, werden verstärkt. Dadurch entstehen, ggf. in mehreren Schritten, überschwellige Verbindungen, die für sich genommen, ohne dass eine bestimmte Kombination von Inputsignalen erforderlich wäre, das Feuern der Zelle auslösen können (vgl. Teil 2, Abschnitt 2.4.7).

Für die an Sequenzelementen für lexikalische Ausdrucksseiten anzunehmenden Lernprozesse gilt, dass sie immer die Spezialisierung auf eine Kombination von Imputerregungen oder auf *eine* bestimmte Imputerregung implizieren: Die Verbindungen mit instanzenbildenden Zellen müssen mindestens spezifisch sein für ein Phonem oder eine Merkmalskombination; instanzenbildende Zellen dürfen nicht auf verschiedene Inputphoneme gleichermaßen reagieren. Die sequenzenbildenden Zellen sollen eine vom Sequenzvorgänger erzeugte Erwartung mit einem tatsächlich erfolgenden Input verknüpfen. Damit ist der Lernprozess für beide Zelltypen durch Verstärkung aktivierter Verbindungen und anschließende Fixierung bei überschwelliger Erregung charakterisiert.

Die Einstellung der Parameter bei den Sequenzverbindungen und die darauf bezogenen Lernvorgänge sind allerdings nicht ganz banal. Für das Endstadium des Lernprozesses an Sequenzverbindungen und die dabei minimal und maximal zu erreichenden Verbindungsgewichte, durch die dann die Höhe der Erwartungspotenziale auf den sequenzenbildenden Zellen bestimmt wird, müssen die folgenden Bedingungen beachtet werden:

1. Das in einer sequenzenbildenden Zelle ausgelöste Erwartungspotenzial muss so hoch sein, dass es trotz des exponentiellen Abklingens den größten zu erwartenden Abstand zwischen aufeinanderfolgenden phonologischen Inputimpulsen überbrückt. Begründung: Die Abarbeitung der Sequenz würde sonst abbrechen.

2. Das in einer sequenzenbildenden Zelle ausgelöste Erwartungspotenzial muss so niedrig sein, dass sein Abklingen abgeschlossen ist, wenn der Inputimpuls des übernächsten Inputelements erscheint. Begründung: Ein Lexikon mit dem Eintrag /maus/ (*Maus*) würde sonst auch (als völlig unauffällige Variante) /mus/ (*muss*) akzeptieren.

Da das Abklingen des EPSP in einer zellspezifischen und nicht flexibel veränderbaren Rate erfolgt und auch nicht angenommen werden kann, dass die Gewichte der sequenzenbildenden Verbindungen von Fall zu Fall an die Geschwindigkeit des sprachlichen Inputs angepasst werden, müssen diese Bedingungen variabel für alle zu erwartenden Geschwindigkeiten erfüllt werden. Das bedeutet, dass der längste Kurzvokalabstand kleiner sein muss als

zwei kürzeste Kurzvokalabstände. Dabei ist der kürzeste Kurzvokalabstand als kürzester möglicher Abstand zwischen zwei von der Phonologie an das Lexikon übergebenen Impulsen definiert.

Die Abbildung 4.3.3–2 veranschaulicht diese Bedingungen.

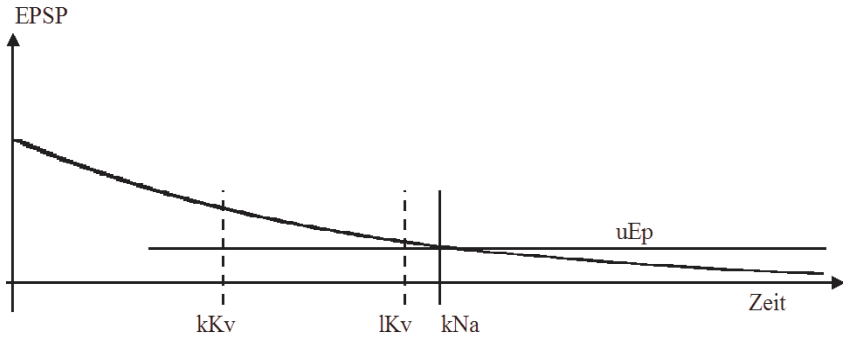

Abbildung 4.3.3–2: Anforderungen an den Verlauf des EPSP in sequenzenbildenden Zellen (Sequenznachfolgern). Der Nullpunkt der Zeitskala entspricht dem Zeitpunkt des Entstehens eines Erwartungspotenzials in einer sequenzenbildenden Zelle. Die Entwicklung des Erwartungspotenzials in der Zeit wird in Beziehung gesetzt zu dem Pegel eines gerade noch unwirksamen Erwartungspotenzials (uEp). Eingetragen sind: der Abfragezeitpunkt, der dem kürzesten Kurzvokalabstand entspricht (kKv), der Abfragezeitpunkt für den längsten Kurzvokalabstand (lKv) und der kürzeste Abstand des zweiten Nachfolger-Inputs (kNa).

Es ist die Kombination von drei Parametern, die für die Einhaltung der skizzierten Bedingungen wesentlich ist: die Anfangshöhe des Erwartungspotenzials in der sequenzenbildenden Zelle (also die Leistungsfähigkeit der seqenzenbildenden Synapsen), die Abnahmerate des Erwartungspotenzials in der sequenzenbildenden Zelle und die Höhe des Abfragepotenzials. Wenn man die Leistungsfähigkeit der sequenzenbildenden Synapsen und die Höhe des Abfragepotenzials auf den Werten festhält, die in den vorangegangenen Simulationen verwendet worden sind, kann durch Verändern der Abnahmerate des Erwartungspotenzials in der Simulation demonstriert werden, was passiert, wenn die genannten Bedingungen nicht erfüllt sind. Die folgenden Simulationen zeigen zunächst die Wirkung einer zu raschen, dann die Wirkung einer zu langsamen Abnahme dieses Potenzials.

Simulation:
Wirkung unterschiedlicher Abnahmeraten des EPSP in sequenzenbildenden Zellen. Die verwendete Lexikonstruktur entspricht Abbildung 4.3.2–3. Der Input ist /bana:ne/
Zu kurzes EPSP. Input in langsamer Sprechweise.
Zu langes EPSP. Input in rascher Sprechweise.

Das Verhalten bei zu kurzem EPSP ist banalerweise zu erwarten: Die Sequenz kann überhaupt nicht bearbeitet werden oder bricht, in Grenzfällen, ab. Bei zu langem EPSP entstehen Mehrdeutigkeiten.

Für die Anforderungen an entsprechende Lernprozesse müssen mindestens zwei Fälle unterschieden werden: Das Verstärken der erforderlichen Sequenzverbindungen in einem einzigen Schritt und das Verstärken in zwei bzw. mehreren Schritten.

Für die Demonstration der Lernprozesse wird im Folgenden durchgängig die in Abbildung 4.3.3–3 dargestellte Architektur verwendet.

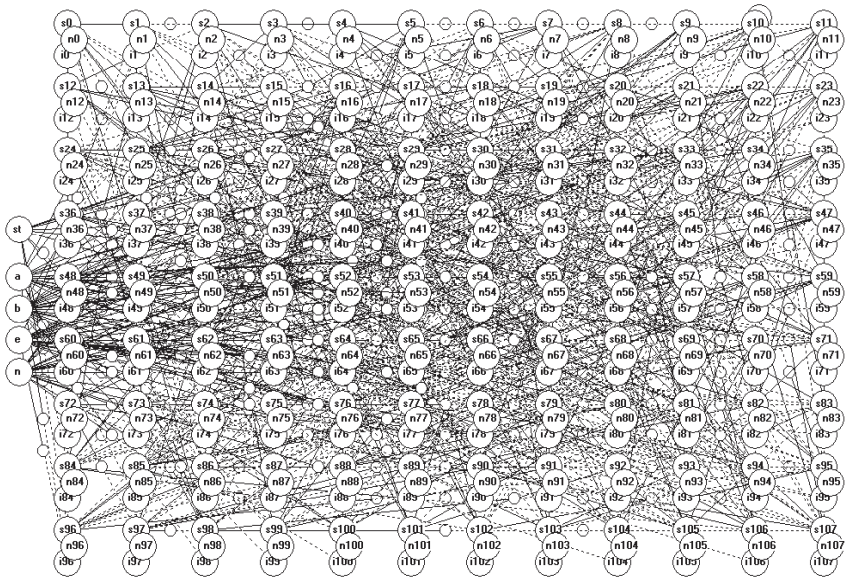

Abbildung 4.3.3–3: Simulationsarchitektur (Bildschirmdarstellung) für lexikalische Lernprozesse.

Die Simulationsarchitektur enthält ein Feld von 108 potenziellen Sequenzelementen aus drei Zellen, die wie oben beschrieben, durch feste Verbindun-

gen verknüpft sind. Das Verbindungsmuster wird auf dem Simulationsbildschirm nicht deutlich und leider läßt es der verfügbare Platz auch nicht zu, die sequenzenbildenden Zellen so darzustellen, dass die Erwartungspotenziale jeweils deutlich sichtbar sind. Am linken Bildschirmrand sind Eingabephoneme zur Darstellung der hier wieder verwendeten Lexikoneinträge /ban/, /ba:n/, /bana:ne/, /an/, /ane/, /a:ne/ und /na:be/ sowie die Startzelle *st* repräsentiert. Die Eingabe bei allen Simulationen besteht in den genannten Wörtern, gefolgt von einer Pause von 27 Sekunden und der Wiederholung derselben Wörter in derselben Reihenfolge zur Kontrolle des Lerneffekts. Bei beiden Listeneingaben variieren die Abstände zwischen den Phonemimpulsen. Während der Wiederholung unterliegt das Netz weiteren Lernvorgängen, das heißt, die Verbindungen sind nicht „eingefroren". Es sind folgende durch Lernvorgänge verstärkbare Verbindungen vorhanden: 10 Verbindungen von *st* zu sequenzenbildenden Zellen, je 25 Verbindungen von den Phonemeinheiten zu instanzenbildenden Zellen, je 7 Verbindungen von den sequenzenbildenden Zellen zu anderen sequenzenbildenden Zellen. Alle diese Verbindungen sind zufällig bestimmt und am Anfang des Simulationsprozesses praktisch wirkungslos, was daran zu erkennen ist, dass die kleinen Kreise, die in der Mitte der Verbindungen jeweils die Verbindungsgewichte symbolisieren, keinen roten Pegel zeigen. In allen Simulationen ist die Lernrate bei den Instanzenzellen so hoch, dass fast immer bei der ersten Erregung schon überschwellige Pegel erreicht werden, was natürlich unrealistisch ist, aber den Simulationsvorgang wesentlich beschleunigt. Am Ende dieses Abschnitts wird dieser Lernprozess noch einmal angesprochen.

Lernen durch Verstärken der Sequenzverbindungen in einem einzigen Schritt

Die Lernrate für die Sequenzverbindungen kann so eingestellt werden, dass eine einzige Benutzung der Verbindung schon das volle (maximale) Gewicht der entsprechenden Synapse erzielt; wenn die betroffene sequenzenbildende Zelle nicht anschließend durch ein Abfragepotenzial zum Feuern gebracht wird, kann das Synapsengewicht vergleichsweise rasch wieder abgebaut werden, sonst wird es längerfristig fixiert. Der Effekt bedeutet ein direktes, sofortiges Einschreiben unbekannter Wörter in den lexikalischen Speicherraum.

Das Ergebnis mit einem ersten Simulationsversuch ist allerdings enttäuschend. Wenn man nach dem ersten Lerndurchgang durch die Eingabeliste und nach der Pause das Lernergebnis mit einem zweiten Durchgang kontrol-

liert, sind Aktionspotenziale an den in Tabelle 4.3.3–1 aufgelisteten Zellen zu beobachten.

/ban/	b/s18 b/s80 b/s99	a/s30 a/s95 **a/s96**	**n/s62** **n/s63** **n/s67** n/s104				
/ba:n/	b/s18 b/s80 b/s99	a/s30 a/s95 **a/s96**	**a/s10** **a/s93**	n/0			
/bana:ne/	b/s18 b/s80 b/s99	a/s30 a/s95 **a/s96**	**n/s62** **n/s63** **n/s67** n/s104	a/s50 a/s56 a/s85	a/0	n/0	e/0
/an/	a/s28 **a/s96**	**n/s63** **n/s67**					
/ane/	a/s28 **a/s96**	**n/s63** **n/s67**	e/s35				
/a:ne/	a/s28 **a/s96**	**a/s10** **a/s93**	n/0	e/0			
/na:be/	n/s48	a/0	a/0	b/0	e/0		

Tabelle 4.3.3–1: Tabelle feuernder sequenzenbildender Zellen bei der Eingabe von Wörtern der Lernliste nach dem ersten Lerndurchgang und der folgenden Pause. Gleichzeitig erregte Zellen sind untereinander gestellt. Es ist jeweils angegeben, welche Phonemgeltung die Zelle hat, gefolgt von der Identifikation der Zelle; n/0 bedeutet, dass keine Instanz für /n/ feuert. Zellen, deren Funktion sich störend auswirkt, sind fett rot ausgezeichnet.

Simulation:
Lernen lexikalischer Ausdrucksseiten I.
Simulation bis Stop führt hinter die Lernphase (Achtung: lange Pause, während der nicht fixierte Verbindungen abgeschwächt werden. Anschließend folgen die Wörter der Eingabeliste zum Test.)

Man erkennt, dass dieselben Zellen an verschiedenen Positionen lexikalischer Ketten auftreten. Es ist zwar zu erwarten und wünschenswert, dass die Kettenanfänge „zusammengelegt“ erscheinen, hier sind aber wesentlich mehr Überlappungen vorhanden. Wenn man mit Hilfe einer Zustandskopie des Netzes, unmittelbar nach der Pause erstellt, eine graphische Darstellung der gebildeten Sequenzverbindungen erzeugt, ergibt sich ein verwirrendes Bild, wie in Abbildung 4.3.3–4 dargestellt.

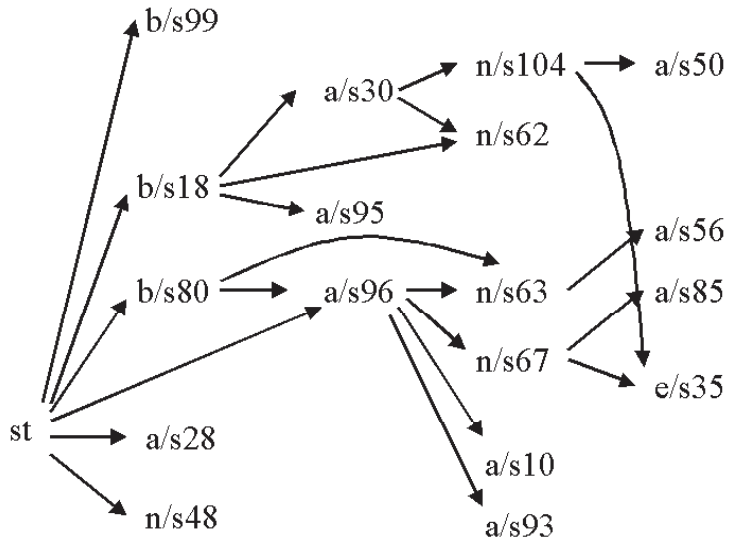

Abbildung 4.3.3–4: Baumdarstellung zu den Werten der Tabelle 4.3.3–1.

Auffällig ist z. B., dass das Sequenzelement mit der sequenzbildenden Zelle *s96*, das für ein /a/ steht, sowohl den Anfang einer der mit /a/ beginnenden Wörter, als auch das zweite Element in einem der mit /ba/ beginnenden Wörter bilden kann. Außerdem wird auf ein beginnendes /b/ unter Auslassung des folgenden /a/ gleich ein /n/ zugelassen. Andere Merkwürdigkeiten sind analog zu interpretieren. In Abbildung 4.3.3–5 ist ein einfaches Beispiel konstruiert, das zeigt, wie diese Phänomene zustande kommen.

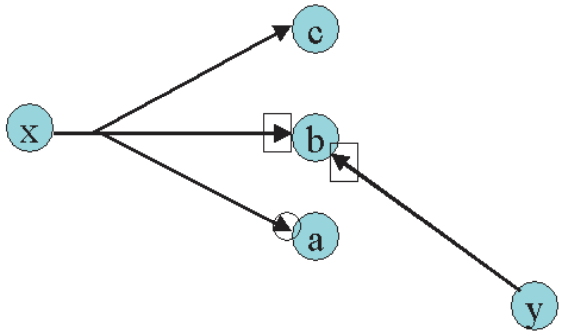

Abbildung 4.3.3–5: Abbildung zur Diskussion von lexikalischen Lernproblemen.

Wir nehmen an, dass x zum Zeitpunkt t1 überschwellig erregt wird und dadurch die Verbindungen mit a, b und c verstärkt werden. Nur die Zelle a ist tatsächlich Nachfolger von x (wie ein entsprechender Input ergibt), die Verbindung von x mit a wird fixiert. Die Zelle y wird zum Zeitpunkt t2 überschwellig erregt und verstärkt eine Verbindung zu b. Die Zelle b ist Nachfolger von y, die Verbindungen von x mit b und y mit b werden durch das Feuern von b fixiert.

Die einfachste Idee zur Behebung des Fehlers ist die Annahme einer Verkürzung der Dauer eines nicht fixierten Synapsengewichts durch Veränderung der Abnahmerate so, dass das Gewicht der Verbindung von x mit b zum Zeitpunkt des Feuerns von b so weit abgeklungen ist, dass die Verbindung unwirksam ist, auch wenn sie fixiert wird.

Die Tabelle 4.3.3–2, die auf dieselbe Weise gewonnen ist wie die Tabelle 4.3.3–1, zeigt, dass diese Maßnahme tatsächlich Erfolg hat.

/ban/	b/s18 b/s80 b/s99	a/s96	n/0				
/ba:n/	b/s18 b/s80 b/s99	a/s96	a/0	n/0			
/bana:ne/	b/s18 b/s80 b/s99	a/s96	n/s63 n/s67	a/0	a/0	n/0	e/0
/an/	a/s28	n/s104					
/ane/	a/s28	n/s104	e/s35				
/a:ne/	a/s28	a/0	n/0	e/0			
/na:be/	n/s48	a/0	a/0	b/0	e/0		

Tabelle 4.3.3–2: Anordnung wie Tabelle 4.3.3–1. Es sind keine störenden Doppelungen zu beobachten.

Simulation:
Lernen lexikalischer Ausdrucksseiten II.
Hinweise zur Simulation siehe Lernen lexikalischer Ausdrucksseiten I.

Eine ergänzende Bemerkung ist am Platz: Da der Lernvorgang sich während der Abfrage der Ergebnisse fortsetzt, ist bei den feuernden Zellen während des Inputs von /bana:ne/ auch das erste /n/ mit zwei Instanzen vertreten. Die entsprechende Baumdarstellung gibt Abbildung 4.3.3–6 :

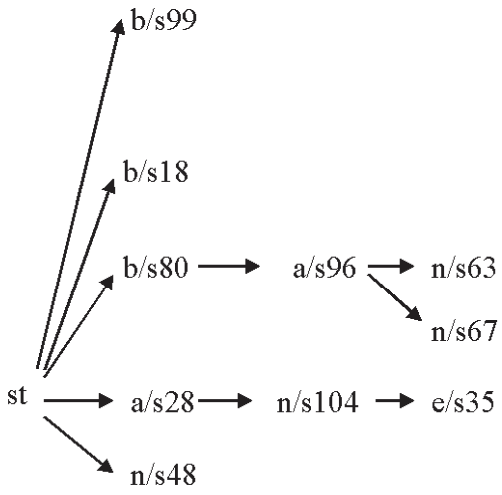

Abbildung 4.3.3–6: Baumdarstellung zu den Werten der Tabelle 4.3.3–2.

Was am Ende bei diesem Simulationsversuch herauskommt, ist deutlich beschränkt durch

- die geringe Zahl von Neuronen und Verbindungen in der Ausgangsarchitektur und
- die, damit verglichen, große Zahl an zu lernenden Lautsequenzen.

Das musste dazu führen, dass einige Sequenzen nicht vollständig repräsentiert werden konnten.

Positiv zu vermerken ist die Redundanz des Ergebnisses: Einzelne Sequenzpositionen werden durch mehrere Sequenzelemente repräsentiert. (Zur Redundanz vgl. Teil 2, „Grundlagen“, Abschnitte 2.3.2 und 2.3.5.) Die Struktur hat, von der Redundanz abgesehen, insgesamt die Form eines Baums, wie im vorigen Abschnitt entwickelt, und mit den dort beschriebenen Verarbeitungseigenschaften.

Andere Eigenschaften des Lernvorgangs und des Lernergebnisses sind diskussionsbedürftig:

- Wenn neue Ausdrucksseiten sofort lexikalisch verankert werden, können unbekannte Ausdrucksseiten nicht mehr als solche erkannt werden, da sie nicht mehr zu Verarbeitungsschwierigkeiten führen. Es ist allerdings

so, dass die Verankerung einer Ausdrucksseite nicht notwendig auch die Verbindung mit einer entsprechenden Inhaltsrepräsentation bedeutet, so dass auf diese Weise unbekannte und bekannte Wörter möglicherweise doch noch differenziert werden könnten.

- Lexikalische Ausdrucksseiten werden im natürlichen Sprachgebrauch mindestens teilweise nicht aus Lautsequenzen abgeleitet, die in Isolation im Input auftreten. Die sofortige Speicherung führt deshalb dazu, dass Sequenzen unbestimmter Länge, das heißt auch u. U. sehr lange Sequenzen, verankert werden. Bei der Natur des sprachlichen Inputs müsste das zur Folge haben, dass ständig eine große Zahl von sehr individuellen Ketten im mentalen Lexikon festgehalten würden, die keine brauchbaren lexikalischen Einheiten darstellen. Das wäre unproblematisch, wenn man mit einem raschen Vergessen solcher Ketten rechnen könnte und mit einem Fortbestehen nur der öfter wiederholten Spuren. Das widerspricht aber der Funktion eines Lexikons, denn dort sollten ja gerade längerfristige Informationen für zukünftige Sprachverarbeitungsvorgänge festgehalten werden. Es ist unklar, wie dieser Widerspruch gelöst werden könnte.

Angesichts solcher Vorbehalte ist es sinnvoll, auch noch die Lernvariante mit kleineren Lernraten und damit mehreren Lernschritten pro zu verankernder Lautsequenz genauer unter die Lupe zu nehmen.

Lernen durch Verstärken der Sequenzverbindungen in zwei oder mehreren Schritten

Das Grundprinzip besteht darin, dass die Wirkung eines Inputs auf die Leistungsfähigkeit einer Sequenzverbindung über eine gewisse Zeitspanne auch ohne Fixierung erhalten bleibt, so dass sich Erhöhungen in zwei oder mehreren Schritten addieren können. Die Festlegung entsprechender Zellparameter ist komplizierter als bei dem Fall des einschrittigen Lernens von Sequenzen. Wenn zwei Inputs erforderlich sein sollen, um eine Sequenzverbindung funktionsfähig zu machen, ergibt sich als zusätzliche Bedingung für die Stärke der Verbindung und die Abnahmerate des Erwartungspotenzials, dass die Wirkung des ersten Inputs ganz (das heißt so weit, dass das Erwartungspotenzial unwirksam wird) abgebaut werden muss, ehe der Nachfolger abgefragt wird. Die sequenzenbildende Zelle der Nachfolgereinheit darf nicht schon im ersten Versuch feuern, die im ersten Versuch erzielte Synapsenstärke darf nicht fixiert werden. Für die Konstruktion des Verhältnisses von Lernrate und Abnahmerate des EPSP ist damit der kürzeste zu

erwartende Abstand zwischen zwei phonologischen Inputimpulsen maßgebend. Andererseits sollte auch der längste Abstand zwischen zwei Inputimpulsen bei zwei(!) Versuchen zu einer Sequenzbildung führen, damit nicht von der Geschwindigkeit abhängt, was gelernt wird. Es wäre wohl nicht sehr plausibel, wenn gerade ein langsamerer Input zu schlechteren Lernergebnissen führen würde. Der Verlauf des EPSP muss entsprechend flach sein. Die Abbildung 4.3.3–7 veranschaulicht diese Bedingungen.

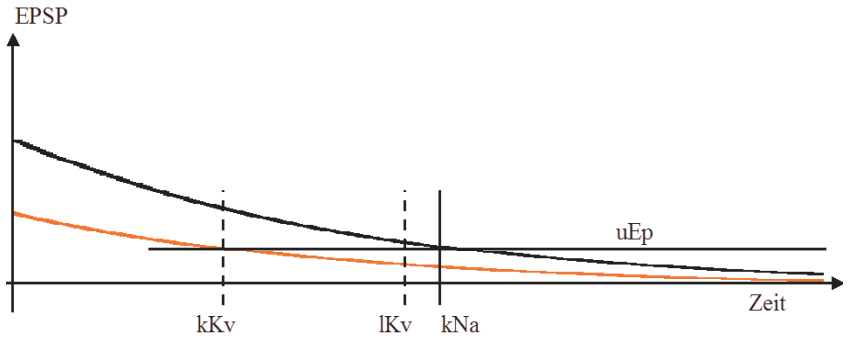

Abbildung 4.3.3–7: Anforderungen an den Verlauf des EPSP in sequenzenbildenden Zellen. Abkürzungen wie in Abbildung 4.3.3–2: uEp = Pegel eines gerade noch unwirksamen Erwartungspotenzials, kKv = Abfragezeitpunkt, der dem kürzesten Kurzvokalabstand entspricht, lKv = Abfragezeitpunkt für den längsten Kurzvokalabstand, kNa = kürzester Abstand des zweiten Nachfolger-Inputs. Die rote (untere) Kurve entspricht dem Potenzial nach dem ersten Lernschritt (unter der Annahme, dass erst zwei Lernschritte eine funktionsfähige Sequenzverbindung aufbauen). Zu dem mit kKv bezeichneten Zeitpunkt darf dieses Potenzial höchstens dem gerade noch unwirksamen Erwartungspotenzial entsprechen.

Wenn man diese Voraussetzungen in der Simulation umsetzt, erhält man wieder unbefriedigende Ergebnisse, die vergleichbar sind mit denen der Tabelle 4.3.3–1. Die Ähnlichkeit der Tabellenwerte spiegelt teilweise das gleichartige Angebot an potenziellen Verbindungen wieder. Die Problemlösung, die oben die verbesserte Tabelle 4.3.3–2 hervorgebracht hat, konnte hier schon allein deshalb nicht übernommen werden, weil sie dem Lernprinzip wiederspricht, das, wie oben bemerkt, das längere Andauern von Lerneffekten an Synapsen voraussetzt.

/ban/	b/s18 b/s80 b/s99	a/s30 a/s95 **a/s96**	**n/s62** **n/s63** **n/s67**				
/ba:n/	b/s18 b/s80 b/s99	a/s30 a/s95 **a/s96**	**a/s10** **a/s93**	**n/s63**			
/bana:ne/	b/s18 b/s80 b/s99	a/s30 a/s95 **a/s96**	**n/s62** **n/s63** **n/s67**	**a/s10** a/s56 a/s85	a/0	n/0	e/0
/an/	a/s28 **a/s96**	**n/s63** **n/s67** n/s104					
/ane/	a/s28 **a/s96**	**n/s63** **n/s67** n/s104	e/s35				
/a:ne/	a/s28 **a/s96**	**a/s10** **a/s93**	n/0	e/0			
/na:be/	n/s48	a/0	a/0	b/0	e/0		

Tabelle 4.3.3–3: Anordnung wie Tabelle 4.3.3–1. Es treten wieder störende Doppelungen auf, die fett rot ausgezeichnet sind.

Simulation:
Lernen lexikalischer Ausdrucksseiten III.
Hinweise zur Simulation siehe Lernen lexikalischer Ausdrucksseiten I.

Die Ursachen für die störenden Phänomene sind die gleichen wie oben für den Fall des einschrittigen Lernens beschrieben. Da es aber hier nicht möglich ist, die erforderliche Begrenzung für die Wirksamkeit der Fixierung von Lerneffekten durch den Trick zu umgehen, dass man die Lerneffekte selbst rasch genug abschwächt, wird ein von der Entwicklung der Synapsengewichte unabhängiger Prozess vorausgesetzt, der eben verhindert, dass eine Fixierung von Verbindungsgewichten erfolgt, deren Verursachung zu weit zurückliegt. Das bedeutet die Einführung eines zusätzlichen synapsenspezifischen Parameters, der als „Fixationszeitfenster“ bezeichnet wird. Wenn dieses Fixationszeitfenster für Sequenzverbindungen unter Beibehaltung der übrigen Definitionen in den gezeigten Simulationen auf 100 Zeittakte (Millisekunden) eingestellt wird, verschwinden die störenden Artefakte und wir bekommen ein Ergebnis wie in Tabelle 4.3.3–4 dargestellt.

/ban/	b/s18 b/s80 b/s99	a/s30 a/s95	n/0				
/ba:n/	b/s18 b/s80 b/s99	a/s30 a/s95	a/0	n/0			
/bana:ne/	b/s18 b/s80 b/s99	a/s30 a/s95	n/0	a/0	a/0	n/0	e/0
/an/	a/s28	n/s104					
/ane/	a/s28	n/s104	e/0				
/a:ne/	a/s28	a/0	n/0	e/0			
/na:be/	n/s48	a/0	a/0	b/0	e/0		

Tabelle 4.3.3–4: Anordnung wie Tabelle 4.3.3–1. Es sind keine störenden Doppelungen zu beobachten.

Simulation:
Lernen lexikalischer Ausdrucksseiten IV.
Hinweise zur Simulation siehe Lernen lexikalischer Ausdrucksseiten I.

Die Einführung eines neuen Parameters ist ein Schritt, der sorgfältig argumentativ gestützt werden sollte. Da ein Typ von Lernvorgängen betroffen ist, der auch für andere Vorgänge, z. B. die einfache Definition von Konzepten zuständig ist, kann man fragen, ob – jetzt technisch gesprochen – die Synapsen auf instanzenbildenden Zellen ähnlichen Bedingungen unterliegen sollten. Das ist tatsächlich der Fall. Der Bedarf tritt dann auf, wenn die Konzeptbildung in mehreren Schritten erfolgt, das heißt, wenn ein Input mehrfach erfolgen muss um zur Fixierung eines Konzepts zu führen. Wenn, wie für Lernvorgänge erforderlich, ein Überangebot von potenziellen Verbindungen vorausgesetzt wird, besteht auch hier die Gefahr, dass eine zurückliegend aktivierte Verbindung, die mit der Repräsentation eines gerade anstehenden Inputmusters nichts zu tun hat, mit in die Fixierung einbezogen wird und die Repräsentation damit in relativ zufälliger Weise stört. Das Fixationszeitfenster für instanzenbildende Zellen ist kürzer anzunehmen als das für sequenzenbildende Zellen und kann etwa 50 Zeittakte betragen.

Es ist hier schließlich noch zu zeigen, dass auch der oben besprochene einschrittige Lernprozess funktioniert, wenn die Kürzung der Dauer der Synapsenverstärkung durch die Einführung des Parameters „Fixationszeitfenster“ ersetzt wird. Die tabellarische Darstellung des Ergebnisses in Tabelle

4.3.3–5 ist nicht ganz identisch mit der von Tabelle 4.3.3–2, zeigt aber keine störenden Überlappungen.

/ban/	b/s18 b/s80 b/s99	a/s30 a/s95 a/s96	n/s67 n/s104				
/ba:n/	b/s18 b/s80 b/s99	a/s30 a/s95 a/s96	a/s10 a/s93	n/0			
/bana:ne/	b/s18 b/s80 b/s99	a/s30 a/s95 a/s96	n/s67 n/s104	a/s50 a/s85	a/0	n/0	e/0
/an/	a/s28	n/0					
/ane/	a/s28	n/0	e/0				
/a:ne/	a/s28	a/0	n/0	e/0			
/na:be/	n/s48	a/0	a/0	b/0	e/0		

Tabelle 4.3.3–5: Anordnung wie Tabelle 4.3.3–1. Es sind keine störenden Doppelungen zu beobachten.

Simulation:
Lernen lexikalischer Ausdrucksseiten V.
Hinweise zur Simulation siehe Lernen lexikalischer Ausdrucksseiten I.

Wenn man den Verlauf des lexikalischen Lernprozesses bei einer Lernrate, die einen mehrschrittigen Lernprozess erfordert, mit dem einschrittigen Verlauf vergleicht, ist der wichtigste Unterschied, dass pro Laut in der Kette die Eingabe eines zusätzlichen Exemplars desselben Worts gebraucht wird; das Lernen von Ausdrucksseiten ist umso langwieriger, je länger das Wort ist. Dieser Effekt wird abgeschwächt durch die baumförmige Struktur des Lernergebnisses, bei der Wortanfänge ja auch aufgrund *verschiedener* Inputwörter etabliert werden können. Eine positive Konsequenz ist, dass Nichtwörter schon auf Grund einer nicht passenden Ausdrucksseite (weil eben der Lernprozess eine Wiederholung voraussetzt) erkennbar werden und dass der zur Verfügung stehende Speicherraum nicht mit nur einmal vorkommenden Ketten ausgefüllt wird.

Abschließende Überlegungen

Die Baumstruktur zum Ergebnis in Tabelle 4.3.3–5 ist, wieder mit Hilfe einer Kopie des Netzzustands nach der Inputpause, in Abbildung 4.3.3–8

wiedergegeben. Sie dient hier nocheinmal zur Veranschaulichung einer typischen lexikalischen Struktur, wie sie durch die angenommen neuronalen Lernprozesse entsteht.

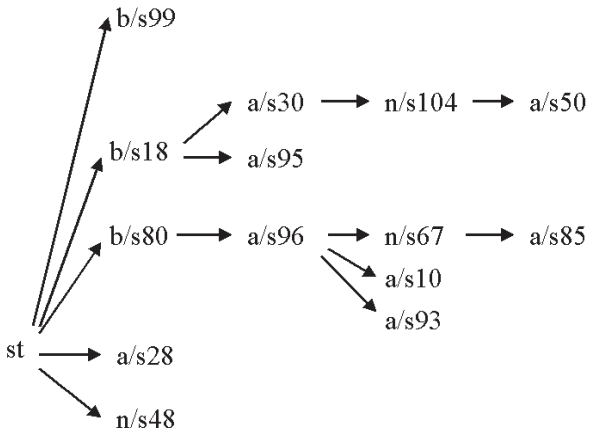

Abbildung 4.3.3–8: Baumstruktur zur Tabelle 4.3.3–5.

Es fällt auf, dass die Lernprodukte der behandelten Strategien, wie das in dieser Abbildung wiedergegebene, keine Verbindungen von sequenzenbildenden Zellen auf sich selbst zurück enthalten. Wenn man die Ausgangsarchitektur nach potenziellen solchen Verbindungen absucht, findet man sie dort auch nur äußerst zufällig, so bei der Zelle *s10*. Nun gilt aber, dass die Möglichkeit der Längung, die diese Verbindungen voraussetzt, für viele Phoneme besteht, nicht nur dort, wo Länge bedeutungsdistinktiv ist. Nicht die Längbarkeit ist die Ausnahme, sondern die Kürze. Man kann auch Zweifel haben, ob die Möglichkeit der Längung von Phonemen und die entsprechende Wahrnehmungsfähigkeit wirklich erst dann gegeben ist, wenn ein Lernprozess stattgefunden hat, der den üblichen lexikalischen Lernprozessen entspricht.

Zu solchen Überlegungen kommt nun hinzu, dass es überhaupt grundsätzliche Schwierigkeiten gibt, einen Lernprozess zu konstruieren, der die schleifenförmigen Verbindungen, auch unter Annahme einer zusätzlichen Zelle, wie in Abbildung 4.3.3–9 angedeutet, verstärken würde. Die Annahme der zusätzlichen Zelle könnte dadurch motiviert werden, dass ja eine Verbindung auf der sequenzenbildenden Zelle nicht mehr etabliert werden kann, wenn diese bereits gefeuert hat. Außerdem wird, um zu verhindern, dass das auf

der Schleife übermittelte Signal in die Refraktärzeit der sequenzenbildenden Zelle fällt, eine gewisse Verzögerung vorausgesetzt, die ebenfalls durch den Umweg über eine zusätzliche Zelle erreicht werden könnte.

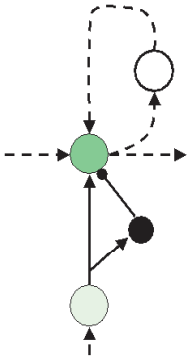

Abbildung 4.3.3–9: Sequenzenbildende Zelle mit potenzieller Wiederholungsschleife und ggf. zusätzlicher Zelle, um Lernprobleme zu umgehen.

Ein entsprechender Lernprozess würde aber auch unter Einbeziehung dieser zusätzlichen Zelle nicht funktionieren, denn er würde ein mindestens zweimaliges Feuern der sequenzenbildenden Zelle im Phonemabstand erfordern, ohne dass die Selbsterwartung bereits etabliert ist. Diese Bedingung ist nun offenbar nicht einzulösen.

Damit muss gelten: Wiederholungsschleifen gehören zur angeborenen Ausstattung sequenzenbildender Zellen. Die Distinktivität von Kürzen kann nur durch einen spezifischen Vergessensprozess gewährleistet werden, nach der Regel:

> Eine Verbindung wird (ggf. in mehreren Schritten) abgebaut, wenn sie nicht innerhalb eines Zeitfensters am Feuern einer Zelle beteiligt ist.

Das Zeitfenster kann identisch sein mit dem Fixationszeitfenster, womit die Regel exakt spiegelbildlich zur Regel für die dauerhafte Fixierung von Verbindungen wird. (Diese Zusammenhänge sind zuerst in Kochendörfer, 2002: 100, formuliert worden.)

Es ist auch noch zu fragen, ob die Alternative einschrittiges vs. mehrschrittiges Lernverfahren eine Entscheidung erfordert. Dazu ist zu bemerken, dass, bei zufälliger Bestimmung von potenziellen Sequenzverbindungen, der Fall eintreten kann, dass die Verbindung zwischen zwei Zellen doppelt vorhanden

ist, so dass, wenn man nicht besondere Begrenzungsmaßnahmen vorsieht, die Wirkung dieser Verbindungen zusammengerechnet schon bei einer ersten Erregung zu Einrichtung einer dauerhaften sequenziellen Abfolge führt, obwohl die Lernraten so eingestellt sind, dass eigentlich eine Wiederholung erforderlich wäre. Die Grenze zwischen den alternativen Verfahren wird durch solche Möglichkeiten verwischt. Man kann dann aber auch grundsätzlich zulassen, dass die Lernraten für potenzielle Sequenzverbindungen variabel sind, vielleicht mit einem relativ großen zahlenmäßigen Übergewicht von Synapsen mit kleineren Lernraten.

Bisher sind in unserer Diskussion und in den Simulationen die an den instanzenbildenden Zellen erforderlichen Lernprozesse vernachlässigt worden. Sie entsprechen den Konzeptbildungsprozessen, die in Teil 2, „Grundlagen“ ausführlich behandelt worden sind. Wenn die Zeitfensterbegrenzung auch für diese Zellen gilt und die Phonologie als Information an das Lexikon nicht Merkmalsbündel sondern Phonemeinheiten liefert, entstehen durch den Lernprozess sozusagen „Kopien“ von Phonemen, die dann in lexikalische Sequenzen eingebunden werden können. Wenn der Lernprozess mehrschrittig ist, ist die Zahl der gebildeten Phonemkopien u. a. auch von der Häufigkeit abhängig, mit der die betreffenden Phoneme im Input vorkommen.

Die beschriebenen lexikalischen Lernprozesse insgesamt liefern redundante baumförmige Strukturen: Jeder Knoten kann mehrfach vorhanden sein, aber das Baumprinzip wird beibehalten, das heißt, dass zwei unterschiedliche Ketten (aus Zellen bestehend, die zu unterschiedlichen Inputketten gehören) nicht im Sequenzverlauf wieder zusammenmünden. Die Überlegungen zu möglichen Verklebungen von Teilstrukturen in lexikalischen Repräsentationen, wie sie in Abschnitt 4.3.2 angestellt worden sind, werden damit durch die Möglichkeit entsprechender Lernvorgänge bestätigt. Es ist jedenfalls nicht möglich, eine listenförmige Lexikonarchitektur zu erzwingen.

4.4 Lexikalische Inhaltsseiten und Modellierung von Produktionsprozessen

4.4.1 Der Weg vom Ausdruck zum Inhalt

Hier sind zunächst einige Bemerkungen am Platz, die Konsequenzen aus allgemeinen Eigenschaften mentaler Repräsentationen betreffen und im Vergleich mit konkurrierenden Modellvorstellungen beachtet werden müssen. Von einem Weg, der vom Ausdruck zum Inhalt im mentalen Lexikon führt, wird gesprochen, weil tatsächlich ein Impulsfluss (kein Informationsfluss, der diese Bezeichnung wirklich verdienen würde) auf einer über Lernvorgänge etablierten bzw. zu etablierenden Bahn angenommen werden muss zwischen den Neuronen, die den Ausdruck und denen, die den Inhalt repräsentieren. Eine bloße Zuordnung, gleich wie man sich das apparativ denken mag, ist nicht möglich, weil sie symbolverarbeitende Prozesse, z. B. eine Adressierung des zugeordneten Elements, voraussetzen würde; man vgl. die allgemeinen Gesichtspunkte in Abschnitt 4.2.1.

Der Endpunkt des Wegs, also der Inhalt, besteht nicht, wie bei verteilten konnektionistischen Modellen, in einem bestimmten Erregungsmuster über mehrere Zellen, das dann als Repräsentation eines Inhalts interpretiert werden müsste, sondern im Prinzip in einer einzelnen Großmutterzelle, die ihre Bedeutung durch ein Netz von Verbindungen erhält, das letztlich seinen Ursprung in der Sinnesperipherie hat. Was sich an einen solchen Endpunkt anschließt, ist nicht die Weitergabe des Inhalts in Form einer kodierten Information, sondern die Produktion eines einzelnen Impulses, der zur Folge hat, dass eine Gedächtnisspur angelegt wird, Vorstellungen produziert werden oder eine motorische Reaktion ausgelöst wird. Die Spezifität des Inhalts wird durch die Auswahl der Zelle bestimmt, die den Inhalt repräsentiert. Das Anlegen einer Gedächtnisspur bedeutet, dass das bestehende Netz durch zusätzliche Verbindungen erweitert wird, die teilweise durch den aktuellen Erwartungskontext mit bedingt sind.

Die Struktur des Netzes, das die Bedeutung einer Großmutterzelle bestimmt, die in dieser Weise als Endpunkt einer von der sprachlichen Ausdrucksseite ausgehenden Verbindung dient, kann sehr variabel sein. Das bedeutet, dass nicht nur Eigenschaften im Sinne klassischer semantischer Merkmale in diesem Netz zeitlich summiert werden, sondern es können auch Strukturen vorhanden sein, die in Teil 2, „Grundlagen“, Abschnitt 2.4.7, als „Szenen“ bezeichnet worden sind.

In Teil 2 ist festgestellt worden, dass die Inhaltsseiten des mentalen Lexikons identisch sind mit außersprachlichen Konzepten und dass die Annahme einer eigenständigen lexikalischen Semantik nicht realistisch ist. Wie die Verbindung der lexikalischen Ausdrucksseiten mit den außersprachlichen Konzepten zu denken ist, ist offen geblieben und muss jetzt genauer spezifiziert werden. Zwei Gesichtspunkte sind dafür zunächst maßgebend:

- Konzepte, die den lexikalischen Ausdrucksseiten entsprechen, entstehen (bzw. werden repräsentiert) unabhängig vom Vorhandensein dieser Ausdrucksseiten.
- Der Vorgang des Sprachverstehens schließt ein, dass diese unabhängig entstandenen Konzepte von den lexikalischen Ausdrucksseiten her aktiviert werden.

Man betrachte nun das Beispiel eines Konzepts, das durch visuelle Wahrnehmungen aktiviert wird. Wenn man eine entsprechende lexikalische Ausdrucksseite befähigen will, dieses Konzept (ersatzweise für die visuelle Wahrnehmung) ebenfalls zu aktivieren, wird eine Verbindung über einen Zelltyp vorausgesetzt, der in Teil 2, Abschnitt 2.4.7, als ODER-Zelle eingeführt worden ist. Die ODER-Zelle verknüpft zwei oder mehrere Inputs so, dass ein Input allein ausreicht, um ein Aktionspotenzial auszulösen. Die für diese Zelle typischen Lernparameter gewährleisten, dass eine Verbindung auch dann noch etabliert werden kann, wenn die Zelle bereits eine Funktion hat. Damit kann das Konzept erworben werden, ehe zusätzlich die sprachliche Ausdrucksseite etabliert wird. Wenn man beachtet, dass gerade komplexere visuelle Konzepte durch Hierarchien von Komponenten definiert sind, aber schließlich im einfachsten Fall in einer einzelnen zusammenfassenden Großmuttereinheit bestehen, liegt es nahe, mit einer Verbindung der lexikalischen Ausdrucksseite zu rechnen, die genau auf diese zusammenfassende Großmuttereinheit führt, wie in Abbildung 4.4.1 dargestellt, weniger mit mehreren Verbindungen, die auf die definierenden Komponenten führen, obwohl diese Möglichkeit nicht auszuschließen ist. Von den Besonderheiten, die durch die allgegenwärtige Redundanz der Repräsentationen entstehen, wird unten in Abschnitt 4.4.2 noch die Rede sein.

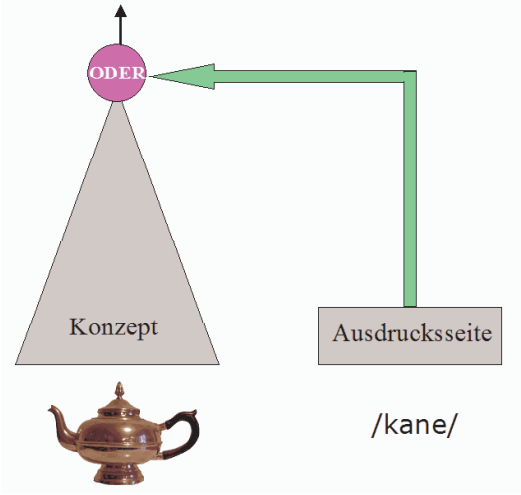

Abbildung 4.4.1–1: Schematische Darstellung der Zuordnung von Ausdruck und Inhalt.

Die lexikalischen Ausdrucksseiten bestehen, wie in den Kapiteln 4.2 und 4.3 gezeigt, aus Ketten von Großmuttereinheiten, wobei diese Großmuttereinheiten jeweils aus einer instanzenbildenden, einer hemmenden und einer sequenzenbildenden Zelle bestehen. Aus den Beobachtungen, die zu dem Kohortenmodell von Marslen-Wilson geführt haben (siehe oben Abschnitt 4.1.2), folgt, dass eine Verbindung, die dem Blockpfeil in Abbildung 4.4.1–1 entspricht, nicht erst vom Ende einer solchen Kette ausgehen kann, sondern dass beliebige Teilketten Verbindungen zu inhaltlichen Konzepten haben müssen. Man könnte daraus die Vorstellung ableiten, dass es eine zusammenfassende Zelle für die lexikalische Kette als Ganzes gibt, mit der jedes einzelne Kettenelement verbunden ist, und die dann ihrerseits eine Verbindung zu dem inhaltlichen Konzept ausbildet. Die zusammenfassende Zelle für die lexikalische Kette müsste aber eine ODER-Zelle sein. Gemäß den in Teil 2, Abschnitt 2.4.7, abgeleiteten Lernbedingungen an ODER-Zellen müsste diese Zelle durch eine überschwellige Aktivation „lernbereit“ gemacht werden. Die einzige Möglichkeit zur Erfüllung dieser Bedingung ist es, eine ODER-Zelle vorzusehen, die mit dem ersten Element der Kette fest und überschwellig verknüpft ist. Mit dieser ODER-Zelle wären dann die nachfolgenden Kettenelemente ebenfalls überschwellig zu verknüpfen. Diese Möglichkeit scheitert aber schon allein daran, dass natürlich alle Kettenelemente vor einem eventuellen Lernprozess „gleichberechtigt“ sein müssen, so dass ODER-Zellen dieser Art bei allen Kettenelementen anzunehmen sind

und damit zwangsläufig die erwünschte zusammenfassende Rolle einer einzelnen Zelle verhindert wird. Weitere Gegenargumente betreffen den Zeitverlauf der Aktivationen an der ODER-Zelle (sehr großer Abstand zwischen den Aktionspotenzialen), sie sollen hier nicht weiter ausgebreitet werden.

Diese Überlegungen führen zu der Annahme, dass die einzelnen Elemente einer lexikalischen Kette nur direkt und ohne eine zusammenfassende Zwischenstufe mit der ODER-Zelle des inhaltlichen Konzepts verbunden sein können, wie in Abbildung 4.4.1–2 dargestellt.

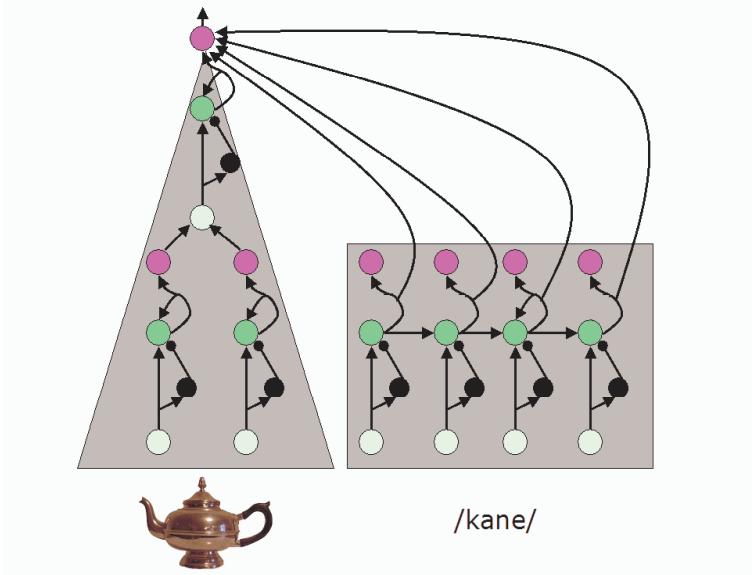

Abbildung 4.4.1–2: Schematische Darstellung der Zuordnung von Ausdruck und Inhalt, Realisierung auf Zellebene. Die Ausstattung einzelner Großmuttereinheiten wird als einheitlich angenommen.

Man beachte, dass die Verbindungen auf die ODER-Zelle des inhaltlichen Konzepts von den *sequenzenbildenden* Zellen der lexikalischen Ausdrucksseite ausgehen. Es ist wegen der Gefahr der Bildung von Endlosschleifen nicht möglich, ODER-Zellen mit ODER-Zellen verknüpft zu denken. Eine ausführlichere Begründung dazu findet sich in Kochendörfer (2002: 109 f.), vgl. auch Teil 2, Abschnitt 2.4.7.

Eine wichtige Konsequenz ergibt sich für den Ablauf des Lernvorgangs: Da die ODER-Zelle lernbereit und damit als Zielzelle der Verbindungen, die von

der Ausdrucksseite her kommen, ausgezeichnet sein muss, ehe diese Verbindungen aktiviert werden, muss immer die Aktivierung des Konzepts der Aktivierung der Ausdrucksseite vorausgehen. Das gilt, genauer, jeweils für jede einzelne der Komponenten der Ausdrucksseite, die ja in zeitlicher Folge erregt werden. Die Erregung des Konzepts muss entsprechend andauern, das heißt zu einer wiederholten Aktivität der beteiligten Neuronen führen.

4.4.2 Lernprozesse

Die Darstellung in Abbildung 4.4.1–2 kann nur als schematisch abkürzend verstanden werden. Eine etwas genauere Analyse muss zusätzlich die folgenden Aspekte einbeziehen:

- Ein Konzept wird, wie oben schon angesprochen, nicht durch eine einzelne Großmuttereinheit, sondern in der Regel durch eine mehr oder weniger große Menge von gleichwertigen Großmuttereinheiten (Instanzen) redundant repräsentiert.
- Ein Konzept erwartet, wenn z. B. die Wahrnehmung eine gewisse Dauer hat, sich selbst und eventuelle Nachfolger. Das ergibt sich selbstverständlich aus der Generalisierung der neuronalen Grundverschaltungen (Plastizität des Gehirns!). Das Konzept selbst kann, technisch, als Nachfolger auftreten (also nicht nur aufgrund von Wiederholungsschleifen, die eine einfache wiederholende Selbsterwartung bewirken).
- Es gibt Ressourcen-Beschränkungen, die vor allem die Menge der möglichen Verbindungen auf den ODER-Zellen betreffen.

Die im Folgenden beschriebene Simulation zeigt zunächst die Besonderheiten, die durch die ersten beiden Punkte, also die Redundanz der Repräsentationen und die Bildung von Instanzenketten, entstehen. Dazu wird ein etwas größeres Netz verwendet, das 1000 Großmuttereinheiten, also insgesamt 4000 Zellen enthält. Der Input geschieht über eine Startzelle, die den entsprechenden Erwartungskontext setzt und insgesamt 4 Zellen, die als Repräsentationen von semantischen Merkmalen interpretiert werden können. Die Einstellung der Lernraten ist so, dass auch ein einzelnes Inputmerkmal eine Instanz bilden kann. Es gibt, von den festen Verbindungen innerhalb der einzelnen Großmuttereinheiten abgesehen,

- je 100 potenzielle Verbindungen auf instanzenbildende Zellen von jeder Merkmals-Input-Zelle aus;

- je 10 potenzielle Verbindungen von sequenzenbildenden Zellen zu anderen sequenzenbildenden Zellen;
- 50 potenzielle Verbindungen von der Startzelle aus auf sequenzenbildende Zellen.

Alle sequenzenbildenden Zellen sind mit Wiederholungsschleifen versehen. Insgesamt sind rund 15 500 potenzielle Verbindungen realisiert. In Abbildung 4.4.2–1 ist ein Simulationsbildschirm dargestellt.

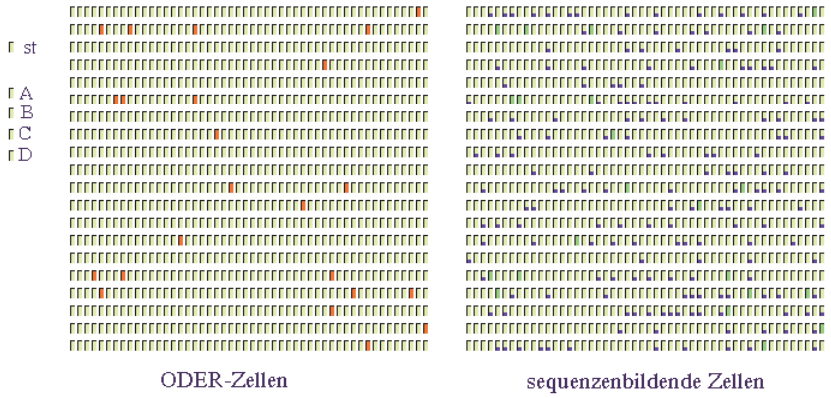

Abbildung 4.4.2–1: Anordnung des Simulationsbildschirms (dargestellt ist der Zustand im Zeittakt 219 der Simulation). Neben bzw. unter den Zellblöcken sind zusätzliche Erläuterungen eingetragen, die nicht auf dem Simulationsbildschirm erscheinen, weitere Details siehe Text.

Um wenigstens ausschnittweise die in dieser Simulation verwendete Menge von Zellen darstellen zu können, wird eine vereinfachte Zelldarstellung verwendet: Jeder Zelle entspricht ein kleines Rechteck. Das Feuern der Zelle wird durch rote, die Refraktärphase durch grüne Ausfüllung des Rechtecks dargestellt. Zusätzlich werden durch verschiedene blaue Pegel noch zwei unterschwellige Potenzialwerte unterschieden: ein in Sequenzverbindungen gerade noch wirksamer niedrigerer Pegel und ein höherer Pegel, der noch unterhalb des Schwellenwerts der Zelle bleibt oder diesen Schwellenwert auch erreicht. Die Verbindungen zwischen den Zellen werden nicht dargestellt. Stattdessen werden einzelne Zelltypen zu Blöcken zusammengefasst. Obwohl es vielleicht verlockend ist, die einzelnen Zellblöcke so zu interpretieren, als würden sie eine Draufsicht auf die Kortexoberfläche wiedergeben, sollte man damit vorsichtig sein, da die Verbindungen zwischen den Zellen ohne Berücksichtigung räumlicher Nachbarschaften geschaltet sind. Ein

Erregungsmuster in einem Zellblock ist also nicht direkt mit dem Ergebnis eines bildgebenden Verfahrens zu vergleichen.

In den in diesem Kapitel folgenden Simulationen wird auf die Bildschirmdarstellung von instanzenbildenden und hemmenden Zellen verzichtet, diese Zellen sind aber vorhanden und normal funktionsfähig. Um die Simulationsverläufe überschaubarer zu halten, sind die Abstände der Inputimpulse immer gleich, von der eigentlich realistischen biologischen Variabilität wird also abgesehen. Aus dem gleichen Grund werden die Komponenten des Merkmalsbündels, das die Instanzen des inhaltlichen Konzepts bildet, gleichzeitig (im gleichen Zeittakt) aktiviert.

Unter diesen Bedingungen ergibt sich in der Simulation sehr drastisch ein Anstieg der Zahl der Instanzen bei jedem Input der Merkmalskomponenten eines Konzepts, sofern eben dieses Konzept (z. B. aufgrund der Wahrnehmung, die dem Konzept zugrunde liegt) länger andauert, das heißt: wiederholt aktiviert wird.

Simulation:
Instanzenbildung bei andauernder Aktivierung eines Konzepts.
Bei langsamen Rechnern kann die Simulation sehr langwierig werden. Die Option „Simulation bis Stop“ sollte deshalb mit Vorsicht verwendet werden.

Die Zahl der gebildeten Instanzen, die am Feuern der Zellen in den beiden auf dem Bildschirm dargestellten Zellblöcken abgelesen werden kann, steigt zunächst annähernd exponentiell, und bleibt dann auf einem Höchstwert konstant, wie in der Grafik der Abbildung 4.4.2–2 veranschaulicht. Das Abbrechen des Anstiegs ist dadurch zu erklären, dass es nur eine bestimmte Zahl von potenziellen Verbindungen der Inputmerkmale mit dem Block der instanzenbildenden Zellen gibt. Wenn die Merkmalszellen jeweils 100 potenzielle Verbindungen mit instanzenbildenden Zellen haben, sind höchstens 400 Instanzen theoretisch möglich, ca. 365 werden tatsächlich gebildet, was auch bedeutet, dass viele davon Instanzen einzelner Merkmale, nicht Instanzen von Merkmalskombinationen sind. Da alle diese Instanzen auch einen aktivierten sequenziellen Kontext voraussetzen, aber nur 50 potenzielle sequenzielle Verbindungen von der Startzelle aus auf den Block der instanzenbildenden Zellen vorhanden sind, muss ein erweiterter Kontext dadurch entstanden sein, dass gebildete Instanzen selbst jeweils sequenzielle Nachfolger erwarten. Die Kontexterweiterung ist in der Simulation an den blauen, länger andauernden Pegeln im Block der sequenzenbildenden Zellen zu erkennen. Die stufenweise Kontexterweiterung führt zu einer wachsenden Ausschöpfung des verfügbaren Instanzenraums.

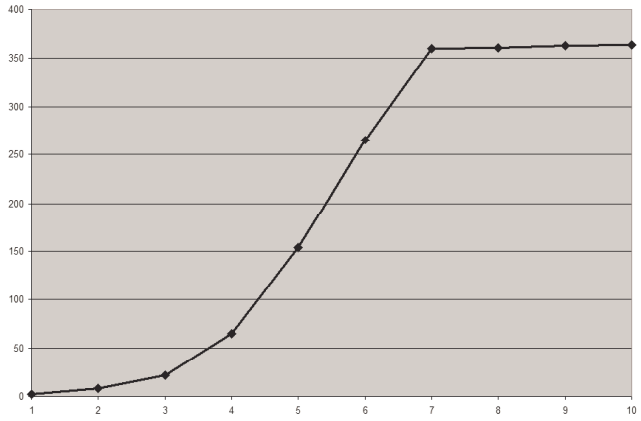

Abbildung 4.4.2–2: Stufenweiser Zuwachs der Zahl der Instanzen für ein Konzept, das durch die andauernde Aktivation von 4 Inputmerkmalen gebildet wird. Die x-Achse zählt die Input-Wiederholungen, die y-Achse gibt die Zahl der Instanzen an.

(Neben den durch den Merkmalsinput direkt ausgelösten Reaktionen in den sequenzenbildenden Zellen entstehen einige Aktionspotenziale auch mit einer geringen Verspätung dadurch, dass die sequenzenbildenden Zellen am Ende sehr eng miteinander verknüpft sind, so dass es passiert, dass so viele *sequenzenbildende* Inputs in eine Zelle vorhanden sind, dass allein, ohne einen zusätzlichen *auslösenden* Input, ein Aktionspotenzial entsteht. Ein Beispiel dafür ist in Zeittakt 500 zu beobachten.)

Die Bedeutung dieser Zunahme der Instanzen kann nun offenbar darin bestehen, dass die ODER-Verbindung der lexikalischen Ausdrucksseiten mit dem inhaltlichen Konzept erleichtert und damit das oben angesprochene Ressourcenproblem abgeschwächt wird, dadurch, dass eine größere Zahl von Anbindungsmöglichkeiten entsteht. Die genauere Analyse dieser Konsequenz ist Gegenstand der nun folgenden Simulationen, in denen versucht wird, eine lexikalische Ausdrucksseite mit dem Ergebnis des Konzeptlernprozesses zu verknüpfen.

Dazu wird zunächst, wie in Abbildung 4.4.2–3 dargestellt, die Simulationsarchitektur um Zellen erweitert, die der Eingabe einer lexikalischen Ausdrucksseite – Phonem für Phonem, in zeitlichen Abständen, wie sie auch im vorangegangenen Kapitel 4.3 verwendet worden sind – dienen.

Es gibt 20 potenzielle Verbindungen von der Startzelle *sta* zu zufällig ausgewählten sequenzenbildenden Zellen und je 100 zufällig ausgewählte Ver-

bindungen von jeder der Zellen *a, b* und *c,* die der Eingabe der phonematischen Einheiten (die nicht weiter spezifiziert werden) dienen, zu instanzenbildenden Zellen. Außerdem sind jeweils 20 potenzielle Verbindungen von jeder sequenzenbildenden Zelle zu ODER-Zellen hinzugefügt. Die Simulation umfasst jetzt insgesamt 35 800 Verbindungen.

Abbildung 4.4.2–3: Architektur zum Lernen von Ausdrucks-Inhaltszuordnungen. Version mit einheitlichen Zellblöcken für Ausdrucksseiten und Inhalte (dieses Charakteristikum wird später variiert).

Der Lernvorgang ist in zwei Phasen unterteilt: Es wird zunächst das inhaltliche Konzept aufgebaut, gefolgt von einer Pause, in der alle Potenziale abklingen, anschließend wird versucht, das aufgebaute Konzept mit einer Ausdrucksseite zu verknüpfen. Wie oben angedeutet, muss die Aktivierung des inhaltlichen Konzepts Phonem für Phonem der Eingabe der Ausdrucksseite vorausgehen.

Simulation:
Lernen einer lexikalischen Ausdrucksseite und ihrer Verbindungen mit Instanzen eines inhaltlichen Konzepts. Wie bei der vorangegangenen Simulation gilt auch hier, dass die Rechenvorgänge bei langsamen Rechnern sehr langwierig werden können. Die Option „Simulation bis Stop“, die direkt an den Beginn der zweiten Phase führt, sollte deshalb ggf. mit Vorsicht verwendet werden.

Wenn man die Simulation über längere Strecken verfolgt, fällt auf, dass wesentlich mehr sequenzenbildende Zellen und ODER-Zellen zum Feuern gebracht werden, als man erwarten würde. Ein krasses Beispiel dafür liefert Zeittakt 810 (Abbildung 4.4.2–4).

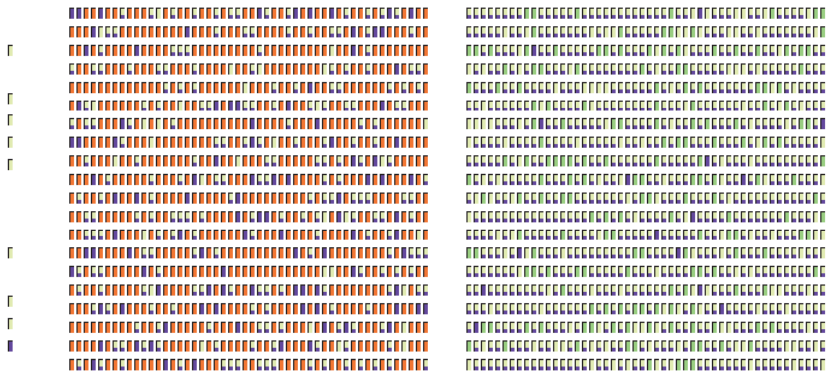

Abbildung 4.4.2–4: Zustand der Simulation zum Erwerb einer Ausdrucksseite und ihrer Verbindungen zu einem inhaltlichen Konzept, mit überschießender Aktivation von ODER-Zellen und sequenzenbildenden Zellen.

Die Ursache für diese Erscheinung wird klarer, wenn man einmal unter Beibehaltung derselben Architektur denjenigen Input des inhaltlichen Konzepts, der unmittelbar vor dem Input des ersten Lauts der lexikalischen Ausdrucksseite liegt, unterdrückt. Der erste Laut löst in diesem Fall gerade noch das Feuern von zwei sequenzenbildenden Zellen aus. Mit der Aktivierung des inhaltlichen Konzepts zusammen sind es aber 19, wobei zu beachten ist, dass der Input des Kontexts über die Startzelle *sta* (siehe oben) maximal 20 feuernde Zellen zulassen würde. Dass davon zufällig insgesamt 19 tatsächlich feuern, ist extrem unwahrscheinlich. Es ist aufgrund dieser Beobachtungen klar, dass das unerwünschte Phänomen durch eine Beeinflussung der Menge der Kontextpotenziale, die für die Ausdrucksseite maßgebend sind, durch die vorangegangene Aktivität des inhaltlichen Konzepts erklärt werden muss. Die Aktivität des inhaltlichen Konzepts hinterlässt eine große Anzahl von Erwartungspotenzialen, die anschließend – fälschlicherweise – durch den lautlichen Input „genutzt“ werden können, obwohl sie mit der Sequenzierung der Ausdrucksseite nichts zu tun haben.

Simulation:
Lernen einer lexikalischen Ausdrucksseite, das inhaltliche Konzept wird vor der Eingabe des ersten Lauts nicht aktiviert.
„Simulation bis Stop“ führt unmittelbar vor die kritische Stelle.
Die Warnhinweise zu den vorangegangenen Simulationen sind auch hier gültig.

Das Entstehen und Stehenbleiben der störenden Erwartungspotenziale kann nicht verhindert werden; die einzige erkennbare Lösung des Problems be-

steht darin, dass man mit einer Bereichstrennung rechnet. Die kontexterweiternden Wirkungen der Aktivation des inhaltlichen Konzepts dürfen den Bereich, der für die ausdrucksseitigen Instanzen vorgesehen ist, nicht erreichen.

Diese Bereichstrennung ist zunächst für die sequenzenbildenden Zellen erforderlich. Bei Licht besehen muss sie aber auch für die instanzenbildenden Zellen gelten, damit unerwünschte Kombinationen von Inhalt und Ausdruck, die bei mehreren Lernschritten auftreten könnten, vermieden werden. Die hemmenden Zellen sind notwendig an die Bereiche der zugehörigen instanzenbildenden und sequenzenbildenden Zellen gebunden. Eine Eigenständigkeit der ODER-Zellen wird man dann mindestens vorläufig nicht annehmen. Sie müssen ja auch feste Verbindungen mit den ihnen jeweils zugeordneten sequenzenbildenden Zellen haben. Selbstverständlich bleibt das Erfordernis erhalten, dass es potenzielle Verbindungen von den sequenzenbildenden Zellen des ausdrucksseitigen Bereichs zu den ODER-Zellen des inhaltsseitigen Bereichs geben muss, die also somit eine bereichsübergreifende Funktion haben. Diskussionswürdig ist, ob umgekehrt auch sequenzenbildende Zellen des inhaltsseitigen Bereichs funktionsfähige Verbindungen mit den ODER-Zellen des ausdrucksseitigen Bereichs aufbauen können. Die folgenden Simulationen schließen eine solche Möglichkeit nicht aus.

Die damit vorausgesetzte Architektur wird in der Bildschirmdarstellung der Abbildung 4.4.2–5 deutlich.

Abbildung 4.4.2–5: Architektur (Bildschirmdarstellung) zum Lexikonerwerb mit Bereichstrennung.

Da die Zahl der Zellen, die für eine bestimmte Verbindungskombination zur Verfügung stehen, sich verändert hat, ist auch die Anzahl der jeweils vorgesehenen potenziellen Verbindungen angepasst worden. So ist die Zahl der potenziellen Sequenzverbindungen zwischen sequenzenbildenden Zellen im Bereich des inhaltlichen Konzepts auf 5 reduziert, im Bereich der Ausdrucksseite auf 15 erhöht worden, letzteres, um noch den kompletten Aufbau der lautlichen Sequenz sicherzustellen. Eine ähnliche Asymmetrie betrifft die potenziellen Verbindungen auf die instanzenbildenden Zellen, deren Zahl für den Inhalt halbiert, für den Ausdruck beibehalten worden ist (100 Verbindungen).

Die Simulation zeigt, dass, wie zu erwarten, die störenden Effekte nicht mehr auftreten. Andererseits ist es jetzt so, dass bei der simulierten Anordnung der erste Laut der Ausdrucksseite nicht mehr mit der Inhaltsseite verknüpft wird, was in der sprachlichen Verarbeitung beim Sprachverstehen zu einer Inkohärenzreaktion führen müsste.

Simulation:
Lernen einer lexikalischen Ausdrucksseite. Das inhaltliche Konzept wird unmittelbar vor Eingabe des ersten Lauts zum ersten Mal aktiviert. „Simulation bis Stop“ führt unmittelbar vor die Lernphase der Ausdrucksseite. Die Warnhinweise zu den vorangegangenen Simulationen bleiben gültig.

Der Rest der Ausdrucksseite wird dann, wie erwartet, mit dem Inhalt verknüpft. Man kann sich überlegen, ob dieses Verhalten als ein häufig auftretender Normalfall akzeptiert werden kann. Letztlich ist aber die konstruierte Situation vielleicht doch etwas künstlich. Wenn man davon ausgeht, dass die Aktivierung des inhaltlichen Konzepts nicht mit der lautlichen Eingabe beginnt und endet, sondern vor allem auch schon *vor* der lautlichen Eingabe erfolgt, verändert sich das Ergebnis positiv.

Simulation:
Lernen einer lexikalischen Ausdrucksseite. Das inhaltliche Konzept wird schon vor der Eingabe des ersten Lauts mehrfach aktiviert. „Simulation bis Stop“ führt unmittelbar vor die Lernphase für die Ausdrucksseite. Abschließend wird zur Kontrolle die Ausdrucksseite allein eingegeben. Die Warnhinweise zu den vorangegangenen Simulationen bleiben gültig.

Bei der abschließenden Eingabe der Ausdruckssequenz allein sieht man, dass beim ersten Laut eine einzige Instanz der Inhaltsseite feuert, beim zweiten Laut sind es 4, beim dritten 11.

Die beschriebenen Lernvorgänge führen zu Konstruktionen, wie in Abbildung 4.4.2–6 abkürzend dargestellt.

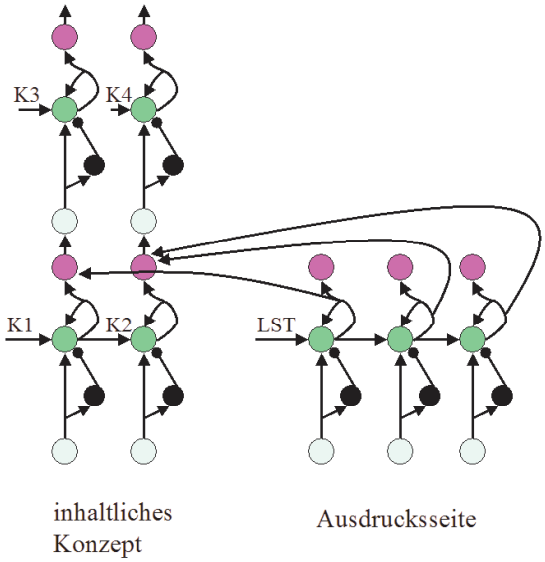

Abbildung 4.4.2–6: Schematische Darstellung einer beim Lernen von lexikalischen Ausdrucksseiten entstehenden Struktur. LST = Lexikonstart, K1 bis K4 Kontextverbindungen für das inhaltliche Konzept. Weitere Erläuterungen im Text.

Die Abbildung zeigt zwei übereinandergestellte Schichten von Großmuttereinheiten. Man muss sich dabei zunächst daran erinnern, dass diese räumliche Anordnung nicht eine Zweischichtigkeit im Kortex wiederspiegelt, sondern dass dort Schichten hierarchischer Strukturen in eine Ebene zusammengeschoben zu denken sind (Teil 2, „Grundlagen“, Abschnitt 2.4.7). Es ist nun interessant, festzustellen, dass bei einem Sprachverstehensprozess die Kontextinformationen (Sequenzverbindungen), die zum Aufbau der Konzeptrepräsentation geführt haben (in der Abbildung 4.4.2–6 K1 und K2) nicht aktiviert zu sein brauchen, so dass die Aktivierung der sequenziell angeschlossenen inhaltlichen Instanz nicht an dem fehlenden Kontext scheitert. Für die obere Zellschicht wird allerdings vorausgesetzt, dass die Kontexte K3 und K4 tatsächlich aktiviert worden sind. Die Verbindungen zwischen der unteren und der oberen Zellschicht müssen aber nicht, wie in der Abbildung, geradlinig und getrennt verlaufen, sondern es sind beliebige Verbindungsmuster möglich. Es sind Weiterverarbeitungsprozesse, die

mit ihren jeweiligen Eigenheiten dafür verantwortlich sind; die Ausdrucks-Inhalts-Zuordnung als solche ist mit der unteren Zellschicht abgeschlossen.

Solange man sich nur auf den Vorgang des Sprachverstehens beschränkt, können Bedingungen und Charakteristika des Lernprozesses für die Zuordnung von lexikalischen Ausdrucksseiten zu Inhaltsseiten jetzt so zusammengefasst werden:

- Die lautlichen Einheiten der lexikalischen Ausdrucksseiten müssen Element für Element mit der ODER-Zelle oder den ODER-Zellen des inhaltlichen Konzepts verknüpft werden.

- Die als mehrfache Instanzenbildung beim Lernen inhaltlicher Konzepte erscheinende Redundanz ist unvermeidlich.

- Wenn man von einem chaotischen Verknüpfungsmuster der potenziellen (lernbaren) Verbindungen auf den ODER-Zellen ausgeht, ist nicht damit zu rechnen, dass alle von *einer* lexikalischen Ausdrucksform ausgehenden Verbindungen auf allen Bedeutungsinstanzen oder auch nur auf einer Bedeutungsinstanz verstärkt werden, sondern es werden nach dem entsprechenden Lernvorgang Verbindungen über mehrere Instanzen verteilt auftreten.

- Der Lernvorgang funktioniert nur dann befriedigend, wenn eine ausreichende funktionale Trennung zwischen den kortikalen Strukturen der Ausdrucksseiten und denen der Inhaltsseiten angenommen werden kann.

- Die Tatsache, dass bei einem bloßen sprachlichen Input die Kontexte *unterhalb* der betroffenen inhaltsseitigen ODER-Zellen fehlen, ist unkritisch, sie werden im Sprachverstehensprozess nicht gebraucht. Das Ressourcen-Problem an den ODER-Zellen wird tatsächlich abgeschwächt.

Die aus den Bedürfnissen des Lernvorgangs abgeleitete Bereichstrennung zwischen inhaltsseitigen und ausdrucksseitigen Strukturen wundert niemanden. Es gehört zu den akzeptierten Erkenntnissen in den Neurowissenschaften, dass unterschiedliche Bereiche des Kortex für unterschiedliche Leistungen zuständig sind. Neu ist allerdings, dass sich zeigen lässt, dass sich diese Eigenschaft zwingend aus den erforderlichen Funktionen ergibt, die letztlich auf – offenbar korrekten – Basisannahmen zu prinzipiellen Eigenschaften neuronaler Repräsentationen beruhen.

4.4.3 Mehrdeutigkeit

Wenn man von Mehrdeutigkeit im lexikalischen Bereich spricht, meint man im Allgemeinen, dass eine lexikalische Ausdrucksseite mit mehreren Inhaltsseiten verknüpft sein kann. Wenn man die Inhaltsseiten identifiziert mit Kategorisierungsmechanismen, die prinzipiell nicht unbedingt sprachlich sein müssen, ergibt sich eine erste Abweichung von dieser Vorstellung, weil dann einige Mehrdeutigkeiten auf Eigenschaften des Kategorisierungsmechanismus zurückgeführt werden können. Das ist eine der Konsequenzen der „Prototypizität“ von Konzepten, wie in Teil 2, „Grundlagen“, Kapitel 2.4 dargestellt. Dabei entsteht eine bemerkenswerte Asymmetrie bezüglich verschiedener sprachverarbeitender Prozesse. In einem rein sprachlichen Verstehensprozess, mindestens für ein isoliertes Einzelwort, tritt diese Mehrdeutigkeit nicht auf, es sei denn, man bezieht Vorstellungsprozesse (vgl. Teil 2, Kapitel 2.5) mit ein. Beim Benennen von Wahrnehmungsobjekten wird sie dagegen relevant. Es ist also, etwas verallgemeinert, ein Unterschied, was die Mehrdeutigkeit im lexikalischen Bereich angeht, wenn man bestimmte Produktionsprozesse einerseits und bestimmte Perzeptionsprozesse andererseits betrachtet.

Es gibt aber selbstverständlich auch Formen von Mehrdeutigkeit, die tatsächlich, wie es der traditionellen Auffassung am ehesten entspricht, in Besonderheiten des Verbindungsmusters zwischen Ausdrucksseite und Inhaltsseite (letztere identifiziert mit dem Kategorisierungsmechanismus) bestehen. Dieser Typ von Mehrdeutigkeit soll hier nun etwas genauer untersucht werden. Oben in Abbildung 4.2.5–1 ist in den Schemata **B** und **C** dargestellt, was sich zunächst als mögliche Differenzierung aufdrängt. Dort ist unterschieden zwischen dem Fall, dass eine Ausdrucksseite mit mehreren Inhaltsseiten verknüpft ist (**B**) oder mehrere gleiche Ausdrucksseiten auf jeweils eins-zu-eins zugeordnete Inhaltsseiten (**C**) führen. Diese naheliegende Alternative wird aber nun sowohl durch die Überlegungen zur Repräsentation der lexikalischen Ausdrucksseiten als auch durch die im vorigen Abschnitt entwickelte Lernproblematik wesentlich verkompliziert. Es kann auch nicht nur darum gehen, wie eine solche Mehrdeutigkeit statisch repräsentiert ist, sondern es muss gleichzeitig beachtet werden, zu welchen Verarbeitungsprozessen sie führt, und damit gehört zu diesem Komplex auch die Frage, ob es Top-down-Verbindungen gibt, die als Rückverbindungen (feedback connections) einen vereindeutigenden Einfluss späterer Verarbeitungsstufen auf frühere ermöglichen (wie z. B. in Frage gestellt von Norris, McQueen & Cutler, 2000; vgl. auch oben 4.2.5).

In den Lernvorgängen des vorangegangenen Abschnitts 4.4.2 ist gezeigt worden, wie einer einzelnen Phonemkette eine Bedeutung zugeordnet werden

kann. Das mentale Lexikon enthält aber, wie in den Abschnitten 4.3.2 und 4.3.3 ausgeführt, nicht notwendig Listen isolierter Phonemketten, sondern komplexere, im Wesentlichen als baumförmig zu charakterisierende Strukturen. Wenn in einer solchen baumförmigen Struktur, bei der dann ein lautliches Element zu mehreren verschiedenen Ausdrucksseiten gehören kann, jede einzelne Ausdrucksseite mit einem Inhalt verknüpft sein soll, ergibt sich die Konsequenz, dass von einem lautlichen Element aus Verbindungen mit mehr oder weniger vielen gleichwertigen oder verschiedenen Inhaltsinstanzen vorhanden sein werden. Letztlich entsteht eine gleichartige Situation aber auch dann, wenn man von einem listenförmigen Lexikon ausgeht, denn der Input eines Phonems wird viele Kettenanfänge gleichzeitig aktivieren, die dann Verbindungen mit den entsprechend vielen Inhalten haben werden. Es ergibt sich ein Phänomen der Mehrdeutigkeit, das nicht Ausdrucksseiten als Ganzes, wie es doch wohl der traditionellen Vorstellung entspricht, sondern einzelne phonologische Elemente betrifft.

Die Mehrdeutigkeit nimmt mit zunehmender Länge der ausdrucksseitigen Kette ab. Es entsteht ein Verlauf, der der Kohortentheorie entspricht, und zwar nicht nur in dem Sinn, dass die Menge der ausdrucksseitigen Kandidaten mit fortschreitendem Verfügbarwerden des lautlichen Inputs, und nur dadurch bedingt, dass Ausdrucksseiten lautliche Kontextketten sind, abnimmt. Es muss mit einer weiteren Wirkung von Kontext gerechnet werden, die das Ergebnis der lautlichen Analyse zusätzlich beschränkt. Diese Kontextwirkung betrifft die Inhalte. Nur bestimmte aus der insgesamt möglichen Menge von Inhalten werden dem inhaltlichen Kontext, in den eine Wortbedeutung eingebettet ist, entsprechen, andere werden durch einen inhaltlichen Kontextfilter ausgeblendet, wie das schematisch in Abbildung 4.4.3–1 wiedergegeben ist.

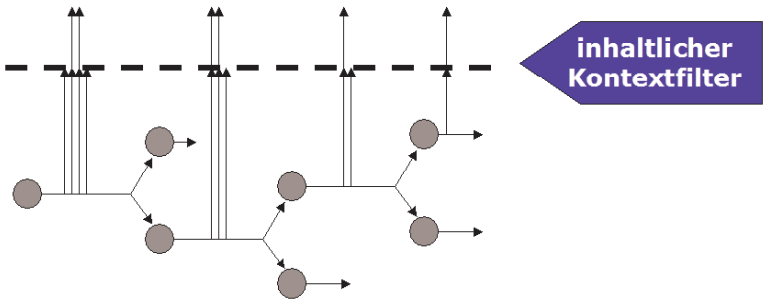

Abbildung 4.4.3–1: Mehrdeutigkeit einer Lexikoneinheit unter Einwirkung eines zusätzlichen inhaltlichen Kontextfilters.

Der inhaltliche Kontextfilter in Verstehensprozessen wird realisiert durch Strukturen, die oberhalb des Kategorisierungskontexts für die lexikalische Bedeutung liegen. Diese Verhältnisse sind in Abbildung 4.4.3–2 in einer auf das Prinzipielle verkürzten Version von Abbildung 4.4.2–6 angedeutet. Man sollte dazu nicht nur den „sequenziellen“ Kontext (Verbindung mit einer sequenzenbildenden Zelle), sondern auch den „punktuellen“ Kontext (Verbindungen mit einer instanzenbildenden Zelle) rechnen. Der ursprüngliche Kategorisierungskontext (in der Abbildung die mit K1 bezeichnete Verbindung) ist am Verstehensprozess nicht beteiligt. Dagegen sind Kontexteinflüsse, die auf noch höherer Verarbeitungsebene angesiedelt sind, selbstverständlich eingeschlossen.

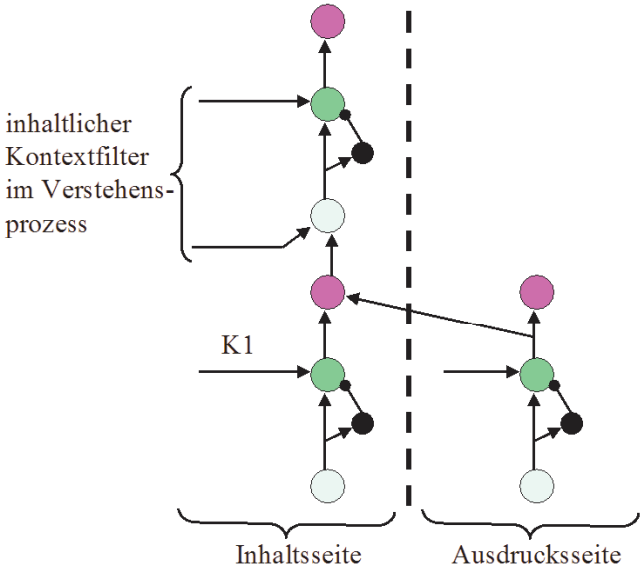

Abbildung 4.4.3–2: Apparative Realisierung des inhaltlichen Kontextfilters. Erläuterungen im Text.

Diese Kontextwirkung, die als reine Filterfunktion gesehen werden kann, impliziert nicht, dass Inhalte auf Ausdrucksseiten zurückwirken. An eine exakte Parallelität von Top-down-Verbindungen über dieselben Zellen, die für die Bottom-up-Verbindungen verwendet werden, mit der Wirkung einer Erregungsverstärkung für bestimmte Zellen, ist schon deshalb nicht zu denken, weil z. B. solche Rückverbindungen von ODER-Zellen auf sequenzenbildende Zellen nicht durch Lernvorgänge verstärkbar wären. Es ist aber

möglich, dass Reparaturphänomene bei Inkohärenz, die technisch identisch sind mit Vorstellungsprozessen (siehe Teil 2, Kapitel 2.4.6), zur Veränderung auch des Kontexts K1 in Abbildung 4.4.3–2 führen. Vorstellungsprozesse, die immer Produktionsprozesse einschließen, können auch die Ausdrucksseiten einbeziehen. Solche Prozesse sind aber nicht selbstverständlicher Bestandteil aller sprachlichen Verstehensprozesse, sondern entstehen in besonderen Situationen. Die Details der lexikalischen Produktion werden in den folgenden Abschnitten behandelt.

4.4.4 Produktionsstrukturen: Grundlagen

In Abschnitt 4.2.4 ist festgestellt worden, dass Produktionsstrukturen in enger Beziehung zu entsprechenden Perzeptionsstrukturen stehen müssen. Das ist schon in Teil 2, Kapitel 2.5, für die Produktion von Vorstellungen gefordert worden und hat zur Konstruktion einer neuronalen Struktur geführt, die in Abbildung 4.4.4–1 wiedergegeben ist. Man beachte, dass „enge Beziehung“ sich zunächst auf die Verbindungsstrukturen bezieht und nicht notwendig(!) eine räumliche Nachbarschaft implizieren muss.

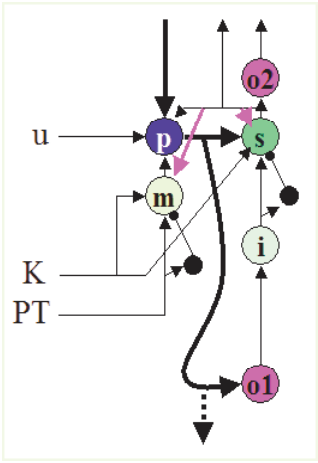

Abbildung 4.4.4–1: Ergänzungen der Perzeptionsarchitektur durch Produktionsstrukturen. Weitere Erläuterungen im Text.

In Abbildung 4.4.4–1 ist rechts ein Element einer Perzeptionsstruktur wiedergegeben, bestehend aus den Zellen *o1, i, s* und *o2*, einschließlich der für die Kontextfunktion erforderlichen Vorwärtshemmung. Die Zellen auf der linken Seite bilden die entsprechende Produktionsstruktur. Die gegenüber

den Perzeptionsbahnen gegenläufigen Produktionsbahnen sind fett hervorgehoben, es sind aber weitere, ergänzende Teilstrukturen vorhanden. Die gesamte Anordnung kann stichwortartig wie folgt charakterisiert werden (eine ausführliche Begründung wird in Teil 2, Kapitel 2.5 gegeben):

- Die mit p bezeichnete Zelle (Typ: „P-Zellen“) ist eine Zelle mit dem Lernmechanismus der ODER-Zellen. Sie ist der eigentliche Kern des Produktionsmechanismus. Verbindungen, die von dieser Zelle ausgehen, können direkt oder über hierarchieniedrigere Zellen gleicher Funktion zur Motorik einerseits und zu einer Zellebene andererseits führen, die als Rückspiegelungsebene bezeichnet worden ist. Die Rückmeldungsebene kann für den Fall der Sprachproduktion, wie in Teil 2, Abschnitt 2.5.3 für das „innere Sprechen“ ausgeführt, mit ODER-Zellen identifiziert werden, die zur Repräsentation phonologischer Merkmale gehören.

  Wichtig sind auch die Verbindungen der P-Zelle mit der sequenzenbildenden Zelle der Perzeptionsstruktur: Eine Produktion führt immer auch zu einer Spur auf der sequenzenbildenden Zelle, die als Erwartungspotenzial in dem zugehörigen Perzeptionsprozess wirkt.

  Die mit u bezeichnete Verbindung gehört zu einem als „unspezifisches prädiktives System“ bezeichneten Verbindungssystem, das für einen normalen, problemlos ablaufenden Produktionsprozess nicht aktiviert zu werden braucht.

- Um eine korrekte Sequenzierung der zu produzierenden Elemente zu gewährleisten, ist ein zusätzliches kontextauswertendes Element erforderlich, das aus der mit m bezeichneten Zelle (Typ „M-Zellen“) und einer Vorwärtshemmung besteht. Die Abfrage dieses Sequenzierungselements wird zentral ausgelöst („Produktionstakt“, in der Abbildung mit PT bezeichnet), das Erwartungspotenzial entsteht durch den sequenziellen Kontext, der identisch ist mit dem Perzeptionskontext (in der Abbildung mit K bezeichnet). Die beteiligten hemmenden Zellen können auch Verbindungen zu mehreren M-Zellen haben, können also anders verschaltet sein, als die im Bereich der Perzeption.

- Alle neu eingeführten Verbindungen sind in ihrer Effektivität vergleichbar mit den strukturell entsprechenden Verbindungen in den bisherigen Architekturen. Die feste Verbindung $s \rightarrow p$ hat Gewicht von 54 % der Schwelle von p, die ebensfalls feste Verbindung von p mit s ein Gewicht von 74% des Schwellenwerts von s. Die Verbindungen zwischen den P-Zellen, die eine Produktionsbahn ausmachen, müssen über Lernprozesse definiert werden, sie können maximal 72% des Schwellenwerts der P-Zellen erreichen.

Aus Elementen der in Abbildung 4.4.4–1 beschriebenen Art kann eine lexikalische Produktionsstruktur aufgebaut werden, wie sie in Abbildung 4.4.4–2 für ein einzelnes Wort dargestellt ist. Die von höheren Verarbeitungsprozessen ausgehende Produktionsbahn verzweigt unterhalb des Elements für die Kombination von Ausdruck und Inhalt auf die einzelnen Elemente der lexikalischen Ausdrucksseite. Während im inhaltlichen Bereich (links) durch die rot eingetragenen Verbindungen eine wiederholte Aktivierung ermöglicht wird, sind diese Verbindungen im Bereich der Ausdrucksseite nicht vorhanden, da alle beteiligten Phoneme Kürzen sind. Die Anbindung der Rückspiegelungsebene an die Struktur der Ausdrucksseiten ist der Übersichtlichkeit halber verkürzt dargestellt, hier sollte es eigentlich eine Zwischenschicht zur Repräsentation von Phonemkonzepten geben, die dann Instanzen zur Einbindung in die Ausdrucksseiten bilden. Das in der vorigen Abbildung mit u bezeichnete unspezifische prädiktive System ist weggelassen, da es zu den dargestellten Prozessen nicht beiträgt.

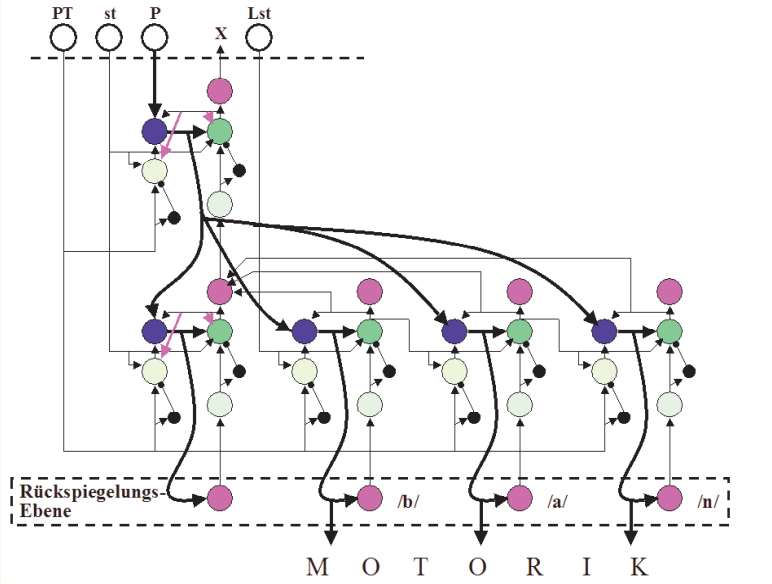

Abbildung 4.4.4–2: Produktionsarchitektur für die Kette /ban/.

Simulation:
Lexikalische Produktion. Der Verlauf kann am besten bei Steuerung über die Leertaste verfolgt werden.
Weitere Erläuterungen im Text.

Der Ablauf der Simulation soll hier etwas genauer beschrieben werden. Um den Nachvollzug zu erleichtern, wird in Abbildung 4.4.4–3 die Bildschirmdarstellung mit den Identifikationen der Zellen wiedergegeben.

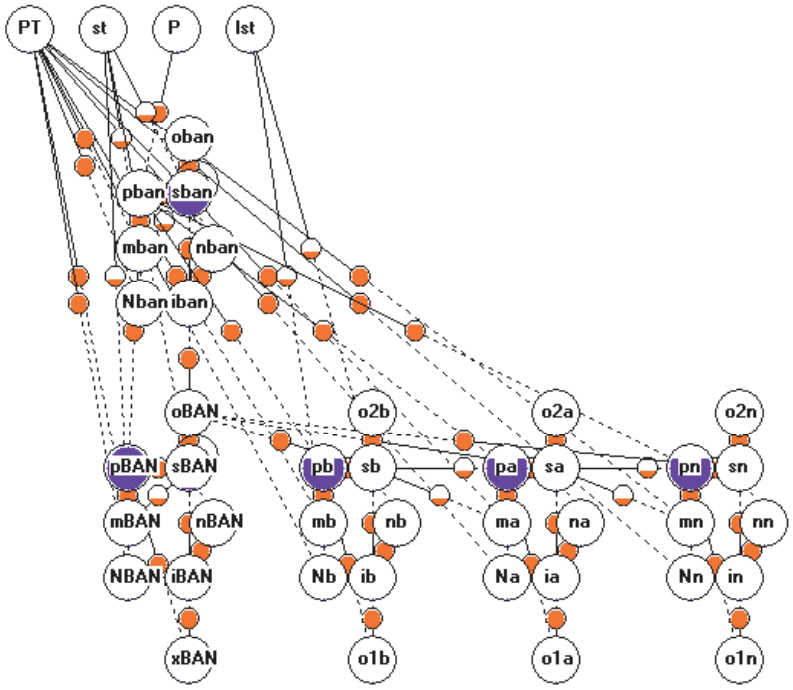

Abbildung 4.4.4–3: Bildschirmdarstellung der Produktionsarchitektur für die Kette /ban/.

Die Simulation beginnt mit dem Setzen eines Kontexts für das inhaltliche Konzept (Zeittakte 10 ff.) und der Aktivierung des inhaltlichen Konzepts, die als ein Vorstellungsprozess verstanden werden kann (Zeittakte 20 ff.). Über den Produktionstakt *PT* und den Mechanismus der M-Zellen werden die zu dem Konzept gehörenden P-Zellen *pban* und *pBAN* mit relativ hohen unterschwelligen Potenzialen versehen. Eine externe Aktivierung von *P*, stellvertretend für entsprechende höhere Prozesse, führt dann zum Feuern der erregten P-Zellen und anschließend zur Rückmeldung, die wie ein Perzeptionsvorgang aussieht. Man beachte die Verzögerung zwischen dem Feuern von *pBAN* und der Ankunft des Aktionspotenzials in *xBAN*, die hier 20 Zeittakte beträgt und im Folgenden als „Rückmeldeverzögerung“ bezeichnet wird. Die Rückmeldung erzeugt Potenziale in den zum Konzept gehörigen P-Zellen, sequenzenbildenden Zellen und M-Zellen, die eine

problemlose Wiederholung des Vorgangs bei entsprechender Aktivität des Produktionstakts ermöglichen. Bemerkenswert ist, dass durch das Feuern von *pban* auch Potenziale auf den P-Zellen der lexikalischen Ausdrucksseite entstanden sind, die allerdings hier ohne besondere Bedeutung sind.

Dann folgt die Produktion der Ausdrucksseite, beginnend mit der Aktivierung des Lexikonstarts (*lst*, Zeittakt 70). Es entsteht dadurch ein Erwartungspotenzial auf *mb* und *sb*. Die Erregung des Produktionstakts (Zeittakt 90) löst jetzt, zusätzlich zum Feuern der M-Zellen der Konzeptrepräsentation, auch das Feuern der Zelle *mb* aus, das zu einem erregenden Potenzial auf *pb* führt. Die eigentliche Produktion durch das Feuern von *pb* ist Folge des Feuerns von *pban*. Der ausdrucksseitige Produktionsprozess besteht also, verallgemeinert, aus zwei Phasen: Es werden zunächst alle sequenziell möglichen Elemente ausgewählt, daran anschließend werden diejenigen Elemente produziert, die zu dem aktuellen inhaltlichen Konzept passen. Wie auch bei der Produktion des Konzepts, erfolgt eine Rückmeldung mit einer Verzögerung von 20 Zeittakten, die einem Perzeptionsprozess entspricht. Erst diese Rückmeldung hat zur Folge, dass über entsprechende Sequenzverbindungen von *sb* auf *sa* und *ma* das nächste Sequenzelement ausgewählt wird. Damit sind die Voraussetzungen geschaffen für die Produktion des nächsten Phonems der Ausdrucksseite. Der Vorgang wiederholt sich jetzt mit jeder Aktivität des Produktionstakts, bis die Ausdruckssequenz vollständig abgearbeitet ist. Analog zum reinen Perzeptionsvorgang gibt auch die lexikalische Produktion jeweils nur einen Impuls pro phonologischer Einheit ab. (Nebenbei sei bemerkt, dass die durch die zeitlich eng aufeinander folgenden Rückmeldungen von Inhaltsseite und Ausdrucksseite entstehende „Doppelwelle“ nicht zum Feuern von *pban* führen darf!)

Die Wahl der das Zellverhalten bestimmenden Parameter ist nicht ganz einfach. So wirkt sich die Länge der Rückmeldeverzögerung auf die Geschwindigkeit aus, die für eine problemlose Produktion gewählt werden kann, wie die Tabelle 4.4.4–1 zeigt. Die beigefügten Simulationen zeigen an Beispielen, mit welcher Art Störungen zu rechnen ist, wenn die entsprechenden Grenzen überschritten werden.

Damit stellt sich die Frage, wie der Produktionstakt überhaupt generiert werden kann. In den bisher gezeigten Simulationen ist er durch einen externen Input erzeugt worden, was natürlich unrealistisch ist. Der externe Input war aber nicht mit anderen Prozessen, abgesehen von einer anfänglichen Kontextsetzung, synchronisiert, so dass man problemlos eine Art innere Uhr annehmen könnte, die den Takt innerhalb der zulässigen Frequenzgrenzen, ohne einen weiteren steuernden Einfluss von dem ausgelösten Geschehen her, generieren würde. Die variable Sprechgeschwindigkeit wird sozusagen

von innen bestimmt oder von Gegebenheiten, die mit der tatsächlichen Produktion nicht direkt etwas zu tun haben.

Abstand der Produktionsimpulse in Zeittakten	Rückmeldeverzögerung 20 Zeittakte	Rückmeldeverzögerung 30 Zeittakte
50	–	–
55	•	–
60	•	•
70	•	•
80	–	•
85	–	•
90	–	–

Tabelle 4.4.4–1: Abhängigkeit der störungsfrei möglichen Produktionsgeschwindigkeiten von der Länge der Rückmeldeverzögerung.

Simulationen:
Typische Produktionsstörungen.
Zu schnell: Rückmeldeverzögerung 20, Produktionstaktabstand 50 Zeittakte.
Zu langsam: Rückmeldeverzögerung 30, Produktionstaktabstand 90 Zeittakte.
Der Verlauf kann am besten bei Steuerung über die Leertaste verfolgt werden.

Eine andere Möglichkeit ist, dass man den Produktionstakt vom Vollzug der Rückmeldung abhängig macht. Der früheste Zeitpunkt, zu dem ein neuer störungsfreier Produktionszyklus eingeleitet werden kann, ist das Erreichen der obersten ODER-Zelle der an der Produktion beteiligten Hierarchie (und damit der Stelle x in der Abbildung 4.4.4–2) durch das Rückmeldungssignal. Wenn man annimmt, dass

- bei der Produktion von Sequenzen immer zwei zeitlich ausreichend eng aufeinanderfolgende Impulse auf der obersten ODER-Zelle produziert werden und dass
- ein reiner Perzeptionsvorgang immer nur einen einzigen Impuls verursacht,

kann man den Produktionstakt direkt aus dieser ODER-Zelle über eine unterschwellige Verbindung ableiten. Im Fall der Produktion addieren sich die Effekte aus den beiden aufeinanderfolgenden Impulsen und lösen das Entstehen eines Produktionstakt-Impulses aus, im Fall eines Perzeptionsvorgangs

wird kein Produktionstakt ausgelöst. Die Produktion muss allerdings immer durch einen unabhängig von der (ja noch nicht erfolgten) Rückmeldung erzeugten Takt *eingeleitet* werden. Bei dieser Konstruktion wird die Produktionsgeschwindigkeit durch die Rückmeldungsverzögerung bestimmt, es ist also eher eine „außen“, nicht “innen“ bedingte Steuerung.

Simulation:
Erzeugung des Produktionstakts über die Rückmeldung.
Die Simulation beginnt mit einer im Verlauf völlig unauffälligen Produktion. Ab Zeittakt 400 folgt ein Perzeptionsvorgang, bei dem der Produktionstakt, wie erwartet, nicht ausgelöst wird.

Beide Möglichkeiten zur Erzeugung des Produktionstakts schließen Mechanismen ein, die außerhalb (oberhalb bzw. unterhalb) der Lexikonstruktur im engeren Sinne liegen. Ein ähnliches Problem zeigt sich, wenn man etwas genauer die Produktion von phonologischen Längen (Langvokalen und längbaren Konsonanten) untersucht. Zu diesem Zweck soll ein erweitertes Modell verwendet werden, das, ergänzend zu dem Lexikoneintrag /ban/ „Bann“ auch noch /ba:n/ „Bahn“ repräsentiert. Die entsprechende Bildschirmdarstellung der Simulation ist in Abbildung 4.4.4–4 gegeben.

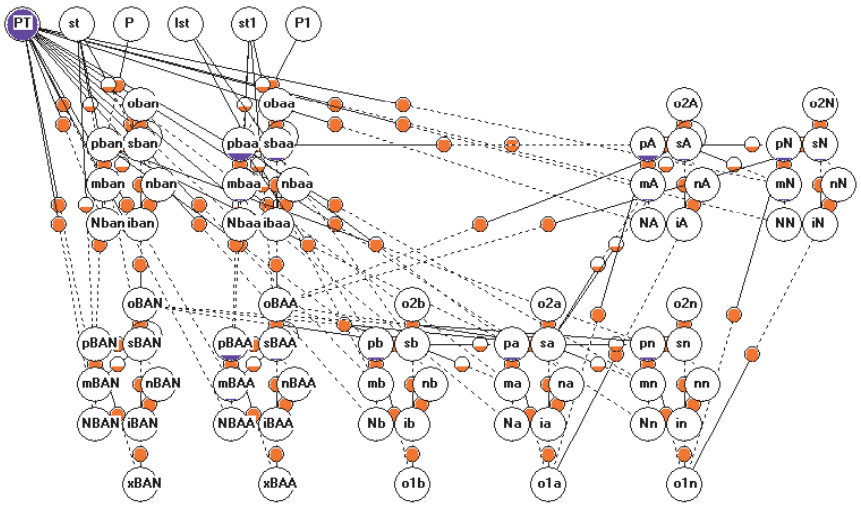

Abbildung 4.4.4–4: Bildschirmdarstellung der Produktionsarchitektur für die Ketten /ban/ und /ba:n/. Gezeigt ist der Simulationszustand im Zeittakt 300, unmittelbar vor der (möglichen) Produktion des /n/ von /ba:n/.

Die Architektur enthält jetzt zwei Konzeptrepräsentationen, die mit den ihnen entsprechenden ausdrucksseitigen Elementen verknüpft sind. Die Struktur der Ausdrucksseiten ist baumförmig, die identischen Wortanfänge (einschließlich des ersten, zur Repräsentation des Langvokals gehörigen Elements) sind „zusammengelegt“. Die auffällige Zerlegung des Langvokals ist aufgrund der Lernprozesse in der Perzeption zwar nicht notwendig zu erwarten, aber andererseits auch nicht grundsätzlich vermeidbar.

Je nachdem, welches der Konzepte zu Beginn der Simulation aktiviert wird, wird entweder /ban/ oder /ba:n/ produziert. Der jeweilige Weg durch das ausdrucksseitige Netz wird durch das Verbindungsmuster der P-Zellen bestimmt, während die Sequenzsteuerung an der Verzweigungsstelle mehrdeutig bleibt. Die ODER-Zellen für die Rückmeldeverbindungen sind für gleiche Laute auch nach der Verzweigung der Ketten identisch.

Die Simulation zeigt, dass die Kette /ban/ ungestört, wie bisher produziert werden kann, die Mehrdeutigkeit der Wortanfänge in der Perzeption ist offenbar kein Problem für die Produktion.

Simulationen:
Produktion von /ban/.
Produktion von /ba:n/.
Der Verlauf kann am besten bei Steuerung über die Leertaste verfolgt werden.

Die Produktion von /ba:n/ führt in Zeittakt 300 zu dem in der Abbildung 4.4.4–4 dargestellten Zustand. Man kann bei genauerem Hinsehen erkennen, dass sowohl mA als auch mN schwache Erwartungspotenziale haben. Wiederholte Produktion wird ermöglicht durch Verbindungen von sA auf sich selbst und auf mA, was erforderlich ist, wenn man von der beliebigen Längbarkeit des Vokals ausgeht. Gleichzeitig wird der Übergang zum Sequenznachfolger vorbereitet, was ebenfalls unvermeidbar ist, da die Beendigung der Vokalartikulation jederzeit erfolgen kann. Konsequenterweise feuern anschließend sowohl pA als auch pN, es ergibt sich also ein Konflikt zwischen den beiden alternativen Lautproduktionen.

Es entstehen in dieser Situation drei Teilprobleme:

(1) Der Konflikt, der durch die gleichzeitige (im biologischen Original: ausreichend gleichzeitige) Aktivierung von Alternativen entsteht, muss zugunsten der einen oder anderen Alternative gelöst werden.

(2) Bei normalem Sprechen ist nur mit einer einmaligen Aktivierung des Längenelements (des zweiten Elements, das den Langvokal repräsen-

tiert) zu rechnen. Die neuronale Architektur muss diese Beschränkung gewährleisten.

(3) Eine willentliche Längung des prinzipiell längbaren Lauts und ein willentlicher Abschluss müssen möglich sein.

Für alle drei Teilprobleme sind zumindest *mögliche* Lösungen vorzuschlagen.

(1) Der Konflikt kann dadurch gelöst werden, dass ein zeitlicher Wettbewerb stattfindet, der bewirkt, dass sich der schnellste (zuerst aktive) Bewerber durchsetzt. Neuronal kann das so realisiert sein, dass der schnellste Bewerber die Konkurrenten hemmt (eine Form lateraler Hemmung). Es wird eine entsprechende Zeitdifferenz zwischen den Aktivationen der Konkurrenten vorausgesetzt.

(2) Eine zufällige Entstehung der in Punkt (1) angesprochenen Zeitdifferenz erklärt den Punkt (2) nicht. Eine gute Möglichkeit ergibt sich aus der Beobachtung, dass bei andauernder Depolarisation von Nervenzellen die entstehende Serie von Aktionspotenzialen abnehmende Frequenz zeigt („Adaptation“, siehe Schmidt, Thews & Lang, 2000: 40 f.). Man wird also annehmen können, dass die folgende Aktivierung einer P-Zelle nicht notwendig mit der gleichen Geschwindigkeit erfolgt, wie eine vorangegangene. Es ergibt sich unter dieser Annahme eine zeitliche Differenz zwischen dem Effekt der *zweiten* Aktivierung der Zelle pA, die später wirksam wird und dem Effekt der *ersten* Aktivierung der Zelle pN, die früher kommt.

(3) Wenn die Aktivierung eines Sequenznachfolgers und der Abbruch einer andauernden Wiederholung des aktuellen Elements durch eine Hemmung dieses Elements geschieht, liegt es nahe, anzunehmen, dass die Wiederholung des aktuellen Elements durch eine Hemmung der Hemmung erzielt werden kann. Gleichzeitig muss aber auch durch eine spezifische Hemmung verhindert werden, dass der Sequenznachfolger des aktuellen Elements ebenfalls feuert. Im Prinzip kann diese Kombination von Leistungen dadurch errreicht werden, dass während eines Produktionszyklus des zu verlängernden Elements nicht nur global die Hemmung ausgeschaltet wird, sondern dass die nicht-aktivierten Elemente eine zusätzliche Hemmung erfahren, die sie daran hindert, in diesem *und* dem folgenden Produktionszyklus aktiviert zu werden. Als wirksames Unterscheidungskriterium zwischen „aktiviert“ und „nicht-aktiviert“ kann die Refraktärphase einer aktivierten Zelle dienen. Die Hemmung bleibt während der Refraktärphase unwirksam. Fällt die zusätzliche Hemmung

in die Refraktärphase der Rückspiegelungszelle des aktivierten Elements, wird dieses Element durch die zusätzliche Hemmung nicht beeinträchtigt. Der Mechanismus, der aus Ausschalten der Hemmung und zusätzlicher Hemmung von Konkurrenten besteht, kann dann so lange getriggert werden, wie die Verlängerung eines Elements gewünscht wird.

Um die etwas komplizierten Zusammenhänge anschaulicher zu machen, sind im folgenden einige Simulationen konstruiert worden, die die erforderlichen Funktionen in typischen Fällen demonstrieren. Dazu dient ein Ausschnitt aus dem Lexikonmodell der Abbildung 4.4.4–4, der die Länge /a:/ in /ba:n/ betrifft und das auf die Länge folgende /n/ (Abbildung 4.4.4–5). Im Unterschied zu den bisher verwendeten Modellen ist eine Phonemebene hinzugefügt, da nicht anzunehmen ist, dass die für ein komplexes Lexikon erforderlichen Lautinstanzen direkt aus phonetischen Merkmalen entstehen können. Die Strukturen dieser Phonemebene sind von denen auf der Ebene der lexikalischen Ausdrucksseiten nicht verschieden, es gibt z. B. Sequenzverbindungen zwischen Phonemen, die dann zustande kommen, wenn eine entsprechende Abfolge im sprachlichen Input öfter auftritt.

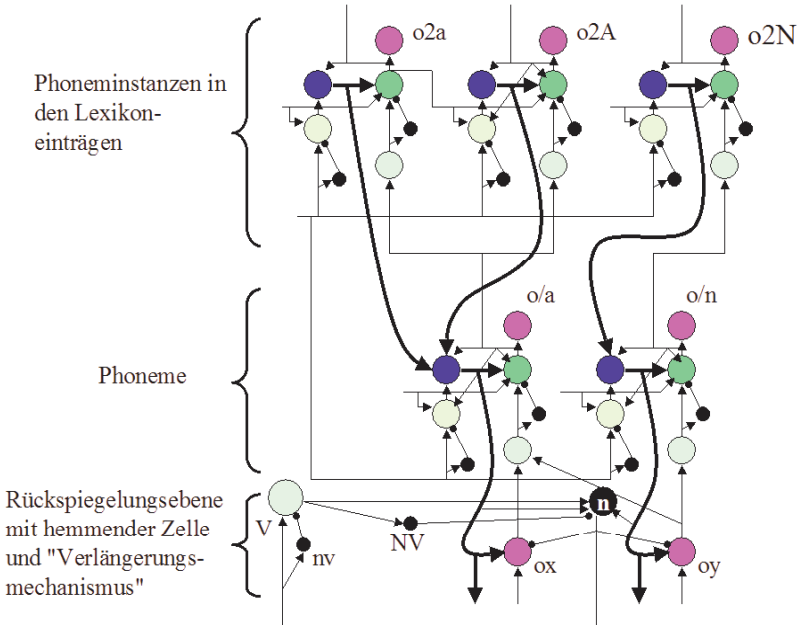

Abbildung 4.4.4–5: Architektur zur Darstellung von Produktionsartefakten und ihrer Verhinderung bei gelängten und längbaren Phonemen.

Die den Aktivierungskonflikt lösende hemmende Zelle ist in der Abbildung mit *n* beschriftet. Sie hat hemmende Verbindungen mit allen Rückspiegelungszellen (hier *ox* und *oy*). Da die Rückspiegelungszellen auf der Ebene der phonetischen Merkmale angesiedelt sind, macht diese Voraussetzung keine Probleme. Wahrscheinlich ist es auch nicht besonders problematisch, von allen P-Zellen, die zu den Phonemen gehören, überschwellig erregende Verbindungen auf diese hemmende Zelle anzunehmen, ggf. können aber statt einer auch mehrere hemmende Zellen dafür zur Verfügung stehen. Die von der Zelle *n* ausgehende Hemmung fällt in die Refraktärphase *erregter* Rückspiegelungszellen und hemmt die übrigen. Dadurch wird bewirkt, dass sich nur das (mit einem kurzen Abstand) zuerst erregte der zur Produktion anstehenden Elemente durchsetzt, das heißt, dass nur dieses zur Rückmeldung kommt. Es entsteht im Fall der Produktion von /a:n/ das normale Verhalten mit nicht zusätzlich gelängtem Langvokal.

Die zusätzliche Längung eines längbaren Vokals wird zunächst durch eine zusätzliche Erregung der Zelle *n* von der Zelle *V* aus bewirkt, deren Hemmung für alle nicht aktiven Zellen das schon vorhandene inhibitorische Potenzial weiter aufstockt, während erregte Rückspiegelungszellen zu diesem Zeitpunkt refraktär und damit unempfindlich sind. Unmittelbar darauf folgend wird dann durch Hemmung der Zelle *n* (auf dem Umweg über *NV*) deren Einfluß auf den folgenden Produktionszyklus ausgeschaltet. Da die Aktivität von *V* eine strenge Synchronisation mit den übrigen Prozessen voraussetzt, liegt es nahe, sie über den üblichen Kontextabfragemechanismus von *n* aus anzusteuern. Das eigentliche die Verlängerung auslösende Signal, das sicherlich keine lexikalische Ursache mehr haben kann, sondern u. U. auch durch eine gleichzeitige außersprachliche Schallwahrnehmung ausgelöst wird, kann dann mit einiger zeitlicher Flexibilität während eines gerade produzierten (und zu verlängernden) Outputs eintreffen. Der Abbruch der Verlängerung wird durch Ausbleiben des Verlängerungssignals bewirkt.

Simulationen:
Produktion der Phonemfolge /a:n/
a) ohne Adaptation,
b) ohne zusätzliche Längung des Vokals,
c) mit zusätzlicher Längung des Vokals.
Man beachte, dass das Simulationsprogramm Potenzialpegel unterhalb des Ruhepotenzials nicht darstellt. Hemmende Effekte werden nur sichtbar an den potenzialsenkenden Konsequenzen, die sie bei Erregung der betroffenen Zellen haben.

Man sollte beachten, dass die zu den bisher verwendeten Grundstrukturen hinzugefügten Zellen nicht zu den Bereichen gehören, in denen Lernprozes-

se stattfinden. Sie können deshalb nicht nur abweichende Verbindungsmuster bilden, sondern auch in ihren lokalen Parametern abweichen. Die Konfliktlösung in der Architektur von Abbildung 4.4.4–5 betrifft allerdings nur die Rückmeldung. Die Bahnen, die zur Motorik führen, zweigen ab, ehe die betreffenden Strukturen erreicht werden. Das bedeutet, dass für die Motorik zusätzlich analoge Strukturen erforderlich sind, die mindestens gewährleisten, dass ein früher ankommender Befehl später ankommende Alternativen hemmt. Wir können nicht mehrere Laute gleichzeitig produzieren und die Entscheidung darüber, was produziert werden soll, kann sicherlich nicht bis in die äußerste Peripherie hinein verschoben werden.

Mit der Erklärung der Produktionstakterzeugung und der Produktion phonologischer Längen ist die Diskussion der Grundlagen lexikalischer Produktion so weit abgeschlossen, dass die Frage nach den erforderlichen Lernprozessen angegangen werden kann. Auf spätere Kapitel verschoben werden die Fragen nach der lexikalischen Repräsentation syntaktischer Informationen (vgl. dazu Teil 5, „Syntax“) und das Silbenproblem (vgl. Abschnitt 4.5). Eine ausführlichere Auseinandersetzung mit den lexikalischen Vorstellungen der Levelt-Schule findet sich in Abschnitt 4.6.

4.4.5 Produktionsstrukturen: Lernprozesse

Man sollte sich an dieser Stelle in Erinnerung rufen, unter welchen Bedingungen zwei Zellen im Kortex durch einen Lernvorgang verbunden werden können:

- Es muss mindestens eine potenzielle (das heißt vor dem Lernprozess etablierte und schwache) Verbindung vorhanden sein, die von der einen Zelle zur anderen führt. Diese Verbindung ist gerichtet, nicht bidirektional, das heißt, Aktionspotenziale bewegen sich nur in immer derselben Richtung.
- Da die Zelle, von der die Verbindung ausgeht, über ihre Axonkollateralen mehrere bis sehr viele Zellen erreicht, muss die Zielzelle durch einen Verarbeitungsprozess herausgehoben sein. Dieser Verarbeitungsprozess kann nicht identisch sein mit der Aktivation der Verbindung (da sonst viele Zellen an die Ausgangszelle angebunden würden), sondern muss eine davon verschiedene Ursache haben.
- Es versteht sich von selbst, dass der Lernprozess auch eine Aktivierung (Erzeugung eines Aktionspotenzials) auf der Zelle voraussetzt, von der die Verbindung ausgeht.

- Der Lernprozess führt zu einer Verstärkung der potenziellen Verbindung, ggf. über eine mehrfache Wiederholung der vorausgesetzten Verarbeitungsprozesse.

In der Abbildung 4.4.5–1 sind alle zur Gewährleistung der Produktion neu zu verstärkenden Verbindungen und die betroffenen (zu erregenden) Zellen gekennzeichnet.

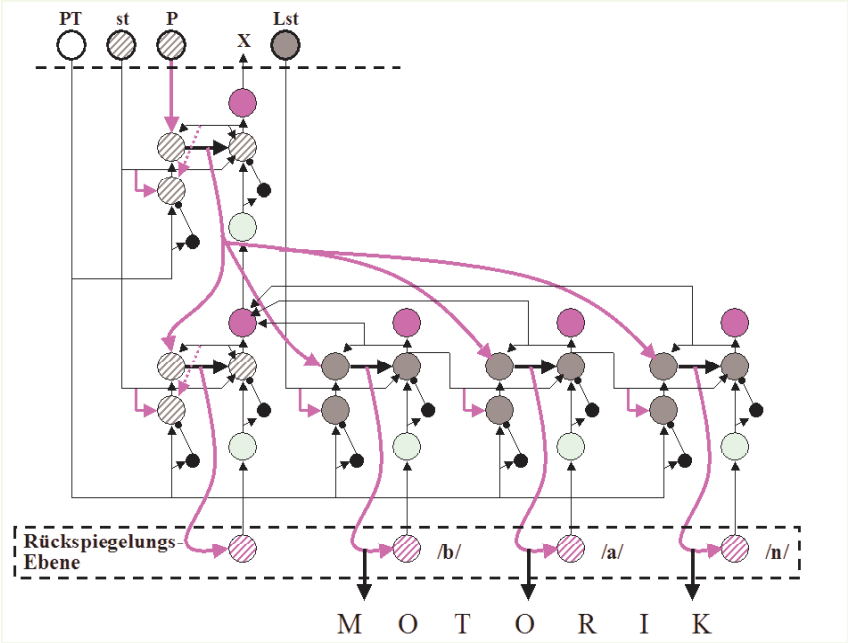

Abbildung 4.4.5–1: Zu verstärkende Verbindungen (rot) und direkt (!) beteiligte Zellen (grau). Weitere Erläuterungen im Text.

Die schraffierten Zellen und ihre Verbindungen sind nicht Gegenstand eines konkreten *lexikalischen* Lernprozesses: Der Aufbau der bedeutungsseitigen Konzepte geht den lexikalischen Prozessen logisch voraus und auch die Rückmeldungsverbindungen sind, wenn man die Strukturen der Abbildung 4.4.4–4 beachtet, nicht Gegenstand einzelner lexikalischer Lernvorgänge. Es bleiben die grau eingefärbten Zellen und die entsprechenden Verbindungen.

Es ist nun nicht möglich, sich ein übergeordnetes System vorzustellen, das genau die in Abbildung 4.4.5–1 grau gefärbten Zellen aktivieren und damit die gewünschten Verbindungen herstellen würde. Es ist ja nicht so, dass

alle Zellen eines bestimmten Zelltyps betroffen sind, sondern ausgewählte Zellen, die jeweils verschiedene lexikalische Informationen repräsentieren. Auswahl und Verbindungsmuster werden durch Lernprozesse bei der Perzeption bestimmt. Das bedeutet, dass ein Perzeptionsvorgang erforderlich ist, wenn entsprechende Produktionsstrukturen gelernt werden sollen. Offenbar genügt es aber nicht, für das Lernen von Produktionsstrukturen nur die Perzeptionsstrukturen zu aktivieren, denn einige der grau gefärbten Zellen – es handelt sich auch dabei um eine informationsspezifische Auswahl – können nur im Rahmen eines Produktionsvorgangs, ausgelöst durch den Produktionstakt, aktiviert werden (vorausgesetzt, dass man nicht zusätzliche parallele Strukturen annehmen möchte).

Aus diesen Überlegungen ergibt sich, dass beim Lernen von Produktionsverbindungen einerseits durch einen entsprechenden Input ein Perzeptionsvorgang stattfinden muss, andererseits muss – in ausreichendem Ausmaß damit koordiniert – der Produktionstakt aktiviert werden. Das Funktionieren dieser Konstruktion kann durch eine Simulation gezeigt werden, bei der Produktionstaktimpulse regelmäßig in den Input eines Perzeptionsvorgangs eingestreut werden (vgl. dazu Kochendörfer, 2002:119). Die Produktionstaktimpulse müssen dabei zeitlich jeweils *nach,* nicht *vor* den zugeordneten Phoneminputs liegen.

Es stellt sich aber auch hier wieder die Frage, woher die Produktionstaktimpulse kommen. In Abschnitt 4.4.4 ist unterschieden worden zwischen einer „inneren“ Steuerung, bei der durch eine „innere Uhr“ Impulse in entsprechenden Abständen erzeugt werden, einerseits, und einer Ableitung solcher Impulse aus dem Rückmeldungsvorgang andererseits, wobei dann der Impulsabstand, also die Sprechgeschwindigkeit, durch „äußere“ Bedingungen bestimmt wird. Die erstere Möglichkeit kann nur dann problemlos angenommen werden, wenn für die Impulsfolge keine weiteren Synchronisationsbedingungen gelten. Das ist aber bei den beschriebenen Lernvorgängen nicht der Fall, der sprachliche Input muss beachtet werden und sein Verlauf ist variabel. Also muss man versuchen, den Produktionstakt, wie oben in Abschnitt 4.4.4 beschrieben, aus der Rückmeldung abzuleiten. Diese Entscheidung gilt dann auch für die einfache Produktion ohne Lernvorgänge. Unter dieser Annahme besteht dann der Input, der dem Lernvorgang zugrunde liegt, nur noch in der Aktivierung des inhaltlichen Konzepts und in unmittelbarem Anschluss daran der Lautsequenz, die von dem Konzept ausgehend nach dem Lernvorgang produziert werden soll.

Damit der Ablauf besser verfolgt werden kann, sind in Abbildung 4.4.5–2 die entsprechenden Symbole, die die Gewichte der zu verstärkenden Verbindungen andeuten, mit Blockpfeilen bezeichnet.

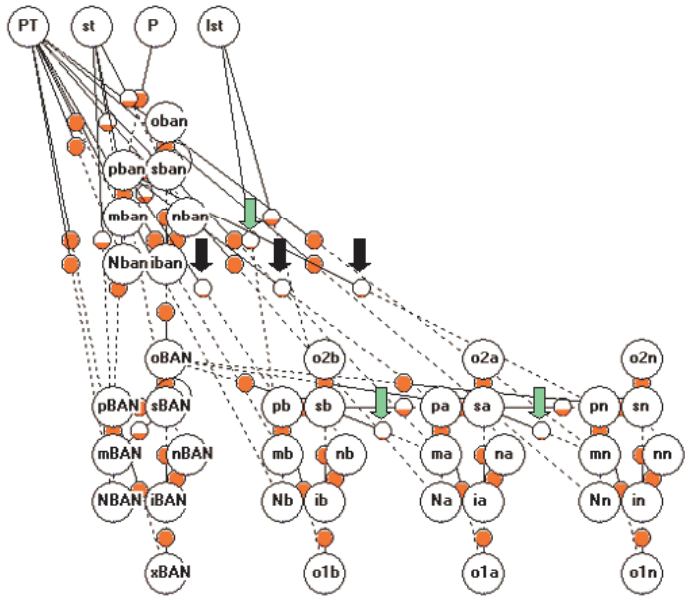

Abbildung 4.4.5–2: Bildschirmansicht mit Hinweisen auf die Darstellungen der sich im Lernprozess verändernden Synapsengewichte. Grüne Pfeile: Sequenzverbindungen, schwarze Pfeile: Verbindungen der Produktionshierarchie.

Simulationen:
Lernen der Produktionsstrukturen für /ban/:
a) innere Steuerung des Produktionstakts,
b) äußere Steuerung des Produktionstakts.

Die bisher besprochenen Modelle zur lexikalischen Produktion gehen von der Annahme aus, dass die lexikalische Ausdrucksseite nicht, wie oben in Abschnitt 4.4.2 beschrieben, mit mehreren inhaltlichen Instanzen verknüpft ist. Man könnte Zweifel daran haben, ob eine Wortform, deren lautliche Elemente Verbindungen mit verschiedenen Instanzen haben, überhaupt störungsfrei produziert werden kann. Die folgende Simulation soll zeigen, dass eine störungsfreie Produktion tatsächlich möglich ist. Die verwendete Architektur entspricht der Anordnung von Abbildung 4.4.2–6, unter Hinzufügung der Rückmeldeverbindungen. Die dort als K3 und K4 bezeichneten Kontexte gelten als gegeben und die Rückmeldung für die inhaltlichen Instanzen,

geht, da es sich um Instanzen *desselben* Konzepts handelt, von ein- und derselben Zelle (in der Simulation *xBAN*) aus. Vgl. dazu auch die folgende Abbildung 4.4.5–3.

Simulation:
Produktion von /ban/ bei Verbindung der Lautkette
mit zwei verschiedenen Instanzen, entsprechend der Abbildung 4.4.2–6.

Schließlich muss jetzt noch gezeigt werden, dass die erforderlichen Produktionsverbindungen, so weit sie nicht fest sind, durch einen unauffälligen Lernprozess erworben werden können. Die Bildschirmdarstellung in Abbildung 4.4.5–3 zeigt, um welche Verbindungen es sich handelt.

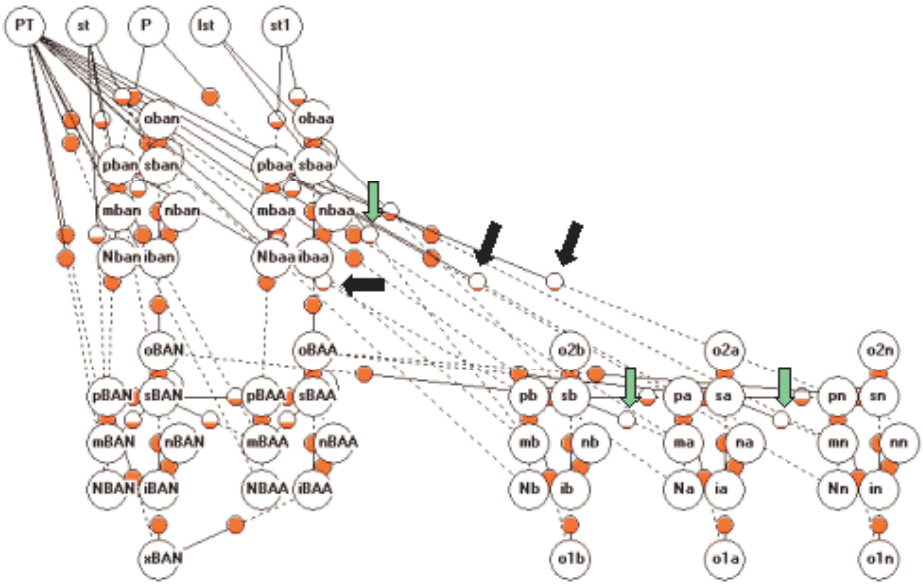

Abbildung 4.4.5–3: Lernvorgang für Produktionsstrukturen bei Verteilung des inhaltlichen Konzepts auf zwei Instanzen: Bildschirmansicht mit Hinweisen auf die Darstellungen der sich im Lernprozess verändernden Synapsengewichte. Grüne Pfeile: Sequenzverbindungen, schwarze Pfeile: Verbindungen der Produktionshierarchie.

Die Simulation verwendet eine „äußere“ Kontrolle des Produktionstakts und ist, wie man am Verlauf sehen kann, ohne Probleme.

Simulation:
Lernen der Produktionsverbindungen von /ban/
unter der Annahme der Verteilung auf mehrere inhaltliche Instanzen.

Eine besondere Schwierigkeit für lexikalische Lernprozesse stellen die Sequenzverbindungen dar, die erforderlich sind, um Langvokale bzw. überhaupt längbare Phoneme zu gewährleisten, also die in Abbildung 4.4.5–4 hervorgehobenen Verbindungen $s \rightarrow s$ und $s \rightarrow m$.

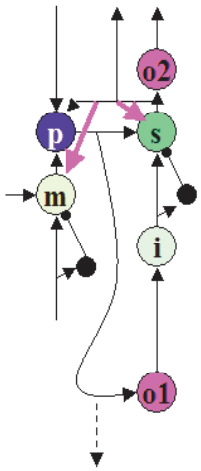

Abbildung 4.4.5–4: Für lange und längbare Phoneme erforderliche zusätzliche Sequenzverbindungen (fett rot eingetragen).

Die Verbindung von s auf sich selbst zurück müsste im Rahmen der ausdrucksseitigen Perzeption aufgebaut werden. Es ist schon oben in Abschnitt 4.3.3 festgestellt worden, dass das nicht möglich ist und dass stattdessen mit einem Vergessensprozess zu rechnen ist, der die funktionsfähig vorausgesetzten Verbindungen dieses Typs, wenn sie nicht gebraucht werden, abschwächt. Die Verbindung von s auf m führt, wie schon in Kochendörfer (2002: 119) festgestellt, auf genau analoge Probleme und muss dementsprechend ebenfalls einem Vergessensprozess unterliegen.

4.5 Silben und Morpheme

4.5.1 Silben in lexikalischen Ausdrucksseiten

Silben sind insofern ein Thema im Rahmen der Behandlung des mentalen Lexikons, als es eine Diskussion darüber gibt, ob silbische Strukturen in irgendeiner Form im Bereich lexikalischer Ausdrucksseiten repräsentiert zu denken sind. Dabei sind zwei Alternativen unterschieden worden:

(1) Die lexikalischen Ausdrucksseiten bestehen nicht aus Einzelphonemen sondern aus Silben, die dann ihrerseits Ketten von Einzelphonemen sind. Die entsprechenden Modelle werden von Roelofs (1997a) als „memory based models“ bezeichnet.

(2) Die lexikalischen Ausdrucksseiten bestehen aus Phonemen, zwischen Lexikon und Phonologie sind aber Regeln und ein Silbeninventar eingeschaltet, die eine silbische Repräsentation gewährleisten. Bei Roelofs (1997a) heißen diese Modelle „rule based models“.

Das Problem ist, soweit prozessorientierte Aspekte eine Rolle spielen, in erster Linie für Sprachproduktionsmodelle diskutiert worden (z. B. Dell, 1986 und 1989; Levelt, 1989; Schade, 1992; Roelofs, 1997a; Levelt, Roelofs & Meyer, 1999; Cholin, Schiller & Levelt 2004). Zur Sprachperzeption gibt es entsprechende Überlegungen z. B. bei Mehler, Dupoux & Segui (1990) und Dupoux (1993).

In den Modellen der Kapitel 4.3 und 4.4 ist weder eine silbische Struktur der Lexikoneinträge noch eine davon zu unterscheidende Silbenebene zwischen Lexikon und Phonologie angenommen worden. Die folgenden Überlegungen sollen zeigen, dass Silbenstrukturen für den lexikalischen Bereich keine Rolle spielen können. Zur Phonologie bzw. Phonetik siehe Teil 3, eine frühere Version ist Kochendörfer (1999a).

Die Alternative (1) ist nur möglich bei der Annahme, dass das mentale Lexikon ein Vollformenlexikon ist, das heißt, dass alle Flexionsformen und Wort-

bildungsprodukte explizit im Lexikon aufgeführt sind und eine morphologische Zerlegung der Ausdrucksseiten nicht vorgesehen ist. Man beachte, dass auch in einer solchen Lexikonstruktur gleiche Wortanfänge „zusammengelegt“ erscheinen können. Wenn das Lexikon kein Vollformenlexikon ist, sind in vielen Fällen während der Sprach*produktion* Prozesse der „Resyllabifizierung“ innerhalb einzelner Wörter (und nicht nur zwischen Wörtern) erforderlich. Das Substantiv *Einheit* muss in *Ein+heit* zerlegt werden. Die Pluralform *Einheiten* in *Ein+hei+ten*, das heißt, das schließende Phonem /t/ wechselt die Silbenzugehörigkeit. Wenn die Morpheme /ainhait/ und /en/ getrennt lexikalisiert sind, ist bei der Bildung der Pluralform eine Änderung der Silbengrenzen erforderlich. Die entsprechenden Prozesse passen aber nicht zu einem speicherbasierten, sondern eher zu einem regelbasierten Modell.

Wenn im Zusammenhang mit der Annahme eines Vollformenlexikons nicht mit Resyllabifizierungen gerechnet werden soll, also Strukturen, die dergleichen leisten können, nicht vorhanden sind, nuss in Kauf genommen werden, dass Resyllabifizierungen auch im Übergang zwischen aufeinanderfolgenden Wörtern (Beispiel bei Roelofs, 1997a, ist die klitisierte Form *ju+we+lin* für *juweel in*) nicht möglich sind. Es sei angemerkt, dass das Lexikon bei Dell (1986) kein Vollformenlexikon ist, das Problem der Resyllabifizierung aber hier und auch in Dell (1988) (vgl. die Diskussion in Roelofs, 1997a) nicht zufriedenstellend behandelt wird.

Die Grundidee einer speicherbasierten Silbenstruktur illustriert Abbildung 4.5.1–1.

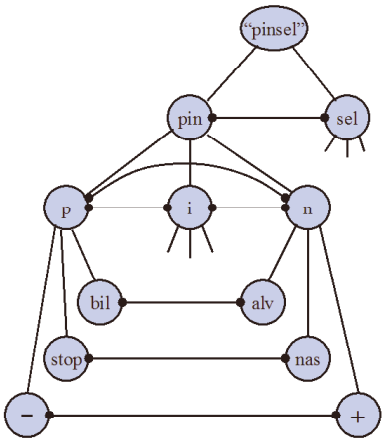

Abbildung 4.5.1–1: Speicherbasierte Silbenstruktur nach Eikmeyer et al., (1991:66); vgl. auch Levelt (1989: 19 und 352).

Die Alternative (2) kombiniert einen Regelsatz und ein Silbeninventar, mit deren Hilfe phonologisch kodierten lexikalischen Ausdrucksseiten eine Silbenstruktur aufgeprägt wird. Die Perspektive ist auch hier die der Sprachproduktion, nicht der Perzeption. Das nach Levelt (2001) vereinfacht wiedergegebene Spreading-activation-Modell der Abbildung 4.5.1–2 soll die Aktivierung und Wirksamkeit des Silbeninventars verdeutlichen.

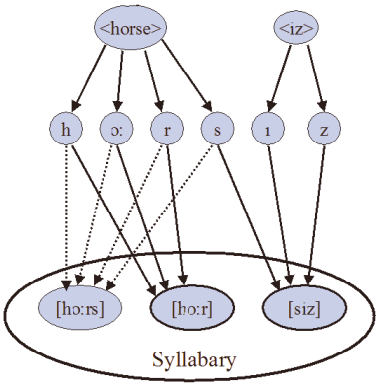

Abbildung 4.5.1–2: Darstellung des Zugriffs auf den Silbenspeicher nach Levelt (2001: Fig. 3).

Der Vorgang der Syllabifizierung wird von Levelt (2001: 13469) so beschrieben:

> „Although we rarely produce a syllable that we never produced before, the low-frequency tail of the distribution may contain syllables which never became stored as motor patterns. The articulation of such items should be prepared on the fly from the syllabified phonological word (ω). Here I will limit to our modeling of syllabary access [...]. It is exemplified in Fig. 3 [siehe Abbildung 4.5.1–2]. As soon as a phonological code is retrieved (for instance the code /hɔrs/), its segments activate all syllabic gestures in the syllabary in which they partake. As an example, the spelled-out segment /s/ spreads activation to the articulatory syllables [hɔrs], [sɪz], and many others. There will, at any one moment, be just one target syllable, given the state of incremental phonological syllabification. Selection of that syllabic gesture occurs under competition, following the same selection mechanism as we saw above for lemma selection. But different from what was the case for lemma selection, the chronometric

> experimental support for the mechanism of syllable selection is still fragmentary.“ (Levelt, 2001: 13469)

Die Rede von einem Wettbewerbsprozess („Selection [...]occurs under competition[...]“) lässt an Vorgänge denken, wie sie in Teil 2, „Grundlagen“, in Abschnitt 2.3.2 für das „competitive learning“ beschrieben sind. An dergleichen einfache, durch laterale Hemmung gekennzeichneten Prozesse ist aber wohl nicht gedacht, wenn man zur Erklärung Roelofs (1997a: 670) heranzieht (worauf die Konzeption bei Levelt, 2001, zurückgeht):

> „For example, WEAVER may syllabify <juweel> with <en> [...]. Following the maximal onset principle in syllabification [...], /l/ will be made onset of the third syllable instead of coda of the second. In this way, WEAVER achieves syllabification across morpheme boundaries and across word boundaries.“

Das darin zitierte „maximal onset principle“ wird in Ewen & van der Hulst (2001: 126 f.) so wiedergegeben:

> „[...] given an intervocalic consonant or consonant cluster, assign it to the beginning of the second syllable, unless an ill-formed syllable-initial cluster is thereby formed. Anything preceeding a well-formed initial cluster is then assigned to the end of the previous syllable.“

Die Einbeziehung eines komplexen Prinzips in die Aktivation eines Netzwerkknotens ist nur dann möglich, wenn ein solcher Knoten eine entsprechend komplizierte Prozedur repräsentiert und nicht nur die einfachen, für ein Spreading-activation-Modell üblicherweise angenommenen Verrechnungen von Aktivationspegeln. Die Interpretation von Netzwerkknoten als Repräsentanten von Prozeduren (die z. B. Etiketten auf den Knotenverbindungen lesen können) gilt generell bei Roelofs (vgl. Roelofs, 1997a: 669), und bringt das Modell in verdächtige Nähe zu prinzipiell unbiologischen symbolverarbeitenden Architekturen.

Die Diskussion der verschiedenen Möglichkeiten zur Repräsentation von Silben im Zusammenhang mit Lexikonstrukturen ist natürlich nur sinnvoll, wenn man überhaupt mit Silben als strukturellen Bestandteilen der sprachlichen Kompetenz rechnen darf. In Ewen & van der Hulst (2001) findet sich ein Kapitel „Why syllables?“, in dem hauptsächliche Argumente *für* eine Rolle des Silbenkonzepts in lexikalischen Strukturen aufgeführt werden. Es wird dort darauf hingewiesen, dass bestimmte intervokalische Konsonantencluster im Wortinneren nur möglich sind, wenn sie eine Silbengrenze

einschließen. Das Argument ist natürlich nur dann stichhaltig, wenn die Silbengrenze nicht gleichzeitig Morphemgrenze ist, und es sich also deshalb verbietet, mit einer ausschließlich morphematischen Segmentierung zu rechnen. Beispiele für das Phänomen, wie etwa im Wort *Bal+drian*, sind im Deutschen relativ selten. Es genügt auch nicht, darauf hinzuweisen, dass Regeln für bestimmte sprachliche Erscheinungen überhaupt oder leichter formuliert werden können, wenn man Silbenstrukturen annimmt. Regelhaftigkeiten im lexikalischen Bereich können sprachgeschichtlich bedingt sein und sich in der Gesamtmenge der Lexikoneinträge wiederfinden, ohne dass entsprechende Regeln tatsächlich als solche repräsentiert zu sein brauchen.

Verschiedentlich ist mit Bezug auf Ergebnisse psycholinguistischer Experimente auch bezweifelt worden, dass Silben in lexikalischen und sublexikalischen Prozessen eine Rolle spielen. Einen Überblick geben Frauenfelder & Floccia (1999: 18 ff., man vgl. auch Content et al., 2001). Ein Problem dabei ist, dass die Silbe als Strukturbestandteil nicht unbedingt universell sein muss, der Hinweis auf Schwierigkeiten damit in bestimmten Sprachen kann nicht entscheidend sein – vorausgesetzt, Silbenstrukturen sind lernbar, das heißt, aus einem sprachlichen Input ableitbar, und damit wenigstens in *einigen* Sprachen real.

Das Problem der Lernbarkeit von Silbenstrukturen wird notorisch vernachlässigt, obwohl sowohl die speicherbasierte Variante (1) als auch die regelbasierte Variante (2) voraussetzen, dass in einem Lernprozess Silben als solche isoliert und entweder in ausdrucksseitige lexikalische Vollformen oder in ein Silbeninventar eingefügt werden können. Weder das spezifische Silbeninventar einer Sprache, noch gar das mentale Lexikon insgesamt können als angeboren (= vererbbar) gelten. Wenn man nun aber genauer untersucht, was ein Lernvorgang für den Erwerb von Silben sein könnte, stößt man auf hartnäckige Widerstände.

Wenn Silben gelernt werden sollen, setzt das zunächst einen input-getriebenen Prozess in Perzeptionsrichtung voraus. Strukturen für die Silben*produktion* können erst in zweiter Linie entstehen. Silben müssen auch perzeptionsseitig *Einheiten* sein, das heißt, sie können nicht einfach als Phonemketten ohne eine zusammenschließende neuronale Einheit aufgefaßt werden. Diese zusammenschließende Einheit muss – in unserer Terminologie – eine ODER-Zelle sein, denn Silben können recht lang werden (Beispiel: *sprichst*) und die Weiterverarbeitung wartet nicht.

Zellen, zwischen denen eine Verbindung gültig verstärkt werden soll, müssen innerhalb eines Zeitfensters ein Aktionspotenzial abgeben. Wenn man zwischen Ausgangszelle und Zielzelle der Verbindung unterscheidet, gilt für Sequenzverbindungen, dass die Aktivität (das Feuern) der Ausgangszelle vor

der Aktivität der Zielzelle liegen muss, bei einer ODER-Verbindung muss es umgekehrt sein. Unabhängig von dieser Reihenfolgebedingung muss aber überlegt werden, wie die „passende“ ODER-Zelle im Fall der Silbenbildung überhaupt aktiviert werden kann. Man beachte, dass diese Zelle silbenspezifisch, das heißt, von Silbe zu Silbe verschieden sein muss. Letzteres bedeutet, dass ihre Aktivierung nur durch den lautlichen Input der Silbe selbst geschehen kann, wenn man die Silbe als rein lautliches Gebilde auffasst. Da Silben prinzipiell Laut*sequenzen* sind, also eine sequenzielle Ordnung gewährleistet werden muss (auch dann, wenn das nicht immer betont und durch die üblichen baumförmigen Darstellungen eher verschleiert wird), ist nicht nur eine zusammenfassende ODER-Zelle erforderlich, sondern auch ein Apparat zur Sequenzrepräsentation. Das ergibt insgesamt, als eine nahe liegende, einfache Architekturidee zur Repräsentation einer Silbe, die Struktur der Abbildung 4.5.1–3.

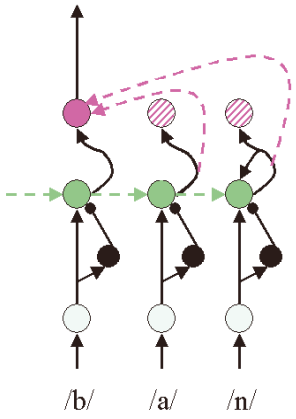

Abbildung 4.5.1–3: Neuronale Struktur einer Silbe. Die durch Lernprozesse zu verstärkenden Verbindungen sind gestrichelt dargestellt. Weitere Erläuterungen im Text.

Hier ist die zusammenfassende Zelle fest mit dem ersten Sequenzelement der Silbe verbunden. Für alle folgenden Sequenzelemente müssen Verbindungen mit dieser Zelle verstärkt werden (rot gestrichelte Linien). Einen Lernprozess setzen auch die grün gestrichelten Sequenzverbindungen voraus. Der Aufbau dieser letzteren Verbindungen ist unproblematisch, er unterliegt den oben in Kapitel 4.3 diskutierten Bedingungen. Die Verbindungen mit der ODER-Zelle können nur verstärkt werden, wenn das Fenster der Lernbereitschaft dieser Zelle, das über die feste Verbindung geöffnet werden kann, ausreichend groß ist, mindestens so groß, wie der Abstand

der Aktionspotenziale auf den zu verknüpfenden Zellen, das ist der Abstand zwischen der Aktivität aufeinanderfolgender Sequenzelemente. Wenn man diesen Minimalabstand annimmt, können aber nur die unmittelbar folgenden Sequenzelemente mit der ODER-Zelle des jeweils vorangegangenen Sequenzelements verknüpft werden. Also könnte man fordern, dass das Lernfenster der beteiligten ODER-Zellen so groß sein muss, dass auch noch das letzte Sequenzelement der Silbe eine Verbindung mit der ersten ODER-Zelle der Silbe aufbauen kann. Diese Idee verbietet sich aber schon allein deshalb, weil Silben sehr verschieden lang werden können und man nicht mit einem Zeitfenster rechnen kann, das an die Länge der erst noch zu lernenden Silbe angepasst ist. Wenn das Zeitfenster länger ist als die zu lernende Silbe, wird die Silbengrenze verwischt, das heißt, eine entsprechende Anzahl von lautlichen Einheiten kann über die eigentlich angestrebte Grenze hinaus der Silbe zugeordnet werden. Das wiederum ist ein Hinweis darauf, dass man nicht nur ein Problem mit dem Silbenende, sondern auch mit dem Silbenanfang hat, vor allem dann, wenn man von einer akustischen Segmentierung des sprachlichen Inputs in Wörter nicht (oder jedenfalls nicht in vollkommener Form) ausgehen kann. Vielleicht ist es möglich, regelhafte Prinzipien zu finden, die doch noch zur Identifikation von Silben führen könnten. Man gerät damit aber, wie schon oben für das Beispiel des „maximal onset principle“ festgestellt, in den Bereich der Möglichkeiten symbolverarbeitender Architekturen. Symbolverarbeitende Architekturen sind aber im Kortex aus vielfachen Gründen nicht anzunehmen.

Das Fazit aus diesen Überlegungen ist, dass die in Abbildung 4.5.1–3 angegebene Silbenstruktur zwar funktionsfähig ist, aber durch Lernprozesse nicht entstehen kann. Wenn man sich überlegt, wodurch diese Schwierigkeit bedingt ist, stellt man fest, dass es die oben schon genannten ganz elementaren Voraussetzungen sind, nämlich dass Silben Einheiten sind und dass die Weiterverarbeitung des sprachlichen Signals das Silbenende nicht abwartet. Unter diesen Voraussetzungen, die so allgemein sind, dass die spezielleren Annahmen über lexikalische Strukturen, die in den vorangegangenen Kapiteln entwickelt worden sind, keine Rolle spielen, können Silbenstrukturen, weil sie nicht lernbar sind, weder in einem speicherbasierten noch in einem regelbasierten Modell der Silbenverarbeitung eine Rolle spielen.

Dieses Ergebnis betrifft zunächst die Sprachperzeption und die dazu erforderlichen Strukturen. Wenn man aber nun hinzunimmt, dass der Aufbau von Produktionsstrukturen für Silben von den entsprechenden Perzeptionsstrukturen abhängig ist, unter der stillschweigenden Voraussetzung, dass man nicht mit Trial-and-Error-Prozessen beim Erwerb von Silbenstrukturen für die Sprachproduktion rechnen kann, wird klar, dass nicht nur die

Perzeptionsstrukturen, sondern notwendig auch die entsprechenden Produktionsstrukturen betroffen sind.

Man beachte jetzt, dass der Hinweis darauf, dass Silbenzerlegungen – mit allerdings vielen bemerkenswerten Unsicherheiten – von linguistisch ungeschulten Sprechern gefunden werden können, kein gutes Argument ist. Es ist nicht notwendig zu erwarten, dass dadurch wirklich interne Verarbeitungsstrukturen gespiegelt werden. Jedenfalls nicht Verarbeitungsstrukturen, bei denen *Repräsentationen* von Silben eine Rolle spielen. Es ist banalerweise nicht zu bezweifeln, dass die Beobachtungen auf Verarbeitungsprozesse zurückgehen, die das entsprechende Verhalten produzieren. Diese Verarbeitungsprozesse dürften aber wesentlich komplexer sein, als nur ein Laut-werden-Lassen von internen Strukturierungen. Man könnte z. B. annehmen, dass Silbenanfänge (beliebiger Länge) bzw. ganze Silben etwas mit möglichen Wortanfängen zu tun haben, dass die Aussprachemöglichkeiten von Plosiven, die Längbarkeit von Konsonanten (Spiranten, Nasale...) eine Rolle spielen und dass schließlich auch die Orthographie und die Silbentrennungsregeln in der Schrift beteiligt sind. Schließlich gilt zweifellos auch, dass man Silbenzerlegungen im episodischen Gedächtnis speichern kann, z. B. als Ertrag des Orthographieunterrichts. Es handelt sich dabei dann aber nicht um Bestandteile der „natürlichen“ Sprachkompetenz, die bei jedem Sprechen zu aktivieren sind, und die Lernbedingungen entsprechen eher denen beim Erwerb morphologischer Einheiten, die im folgenden Abschnitt analysiert werden.

Man kann also zusammenfassen:

- Es genügt nicht, zu zeigen, dass bestimmte Strukturen die Formulierung von Regeln begünstigen, um zu begründen, dass es diese Strukturen und die dazugehörigen Regeln in der Sprachrepräsentation des Kortex (der kein symbolverarbeitender Apparat ist) wirklich gibt.
- Es genügt nicht, zu zeigen, dass (erwachsene?) Sprecher in der Lage sind, einen Ausdruck ihrer Sprache einigermaßen brauchbar in Silben zu zerlegen, solange man nicht geklärt hat, was bei einem solchen Zerlegungsprozess wirklich passiert. Man wird nicht annehmen, dass ein Sprecher einen Einblick haben kann in die mentale Struktur seines Sprachbesitzes.
- Silben sind offenbar nicht lernbar. Lernen von Silben meint hier den Aufbau entsprechender Strukturen *innerhalb* der sprachlichen Kompetenz im ungesteuerten Spracherwerb. Damit können Silben keine strukturellen Komponenten der Sprachrepräsentation im Kortex sein. Das gilt sowohl für die Perzeptionstrukturen als auch für die davon notwendig abhängigen Produktionsstrukturen.

Selbstverständlich ist es eine Konsequenz dieser Überlegungen, dass „Silbeneffekte“ in psycholinguistischen Experimenten nun nicht mehr dadurch erklärt werden können, dass Silbenrepräsentationen für sie verantwortlich sind. Vielmehr ist zu erwarten, dass solche Erklärungen teilweise im engeren Bereich der Phonologie bzw. Phonetik gesucht werden müssen, also unterhalb einer Ebene, die sonst als silbisch angenommen worden ist, teilweise oberhalb dieser Ebene, im Bereich nicht silbisch gegliederter lexikalischer Ausdrucksseiten.

4.5.2 Lernen von Morphemen: Segmentierung

Wenn man, der amerikanischen Tradition folgend, „Morphem“ als Bezeichnung für eine kleinste (kürzeste, gemessen an der Zahl der Phoneme) bedeutungstragende Einheit verwendet, sind Morpheme häufig identisch mit Wörtern und die Lernvorgänge entsprechen denen beim Erwerb von Kombinationen lexikalischer Ausdrucksseiten und Inhaltsseiten. Wenn man die Struktur eines Morphems analog zur Struktur einer Silbe (wie in Abbildung 4.5.1–3) darstellt, bekommt man eine Anordnung wie in Abbildung 4.5.2–1, die identisch ist mit der prinzipiellen Struktur eines Lexikoneintrags, wie in Abbildung 4.4.1–2 skizziert.

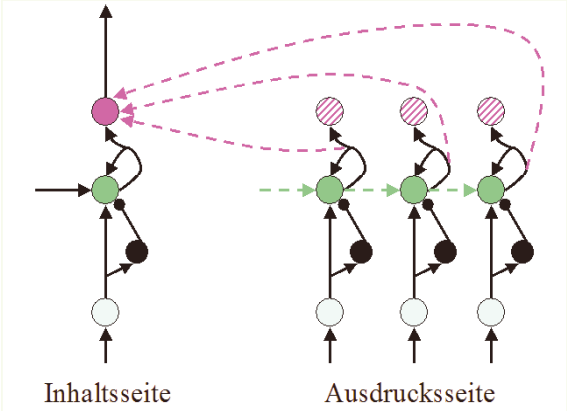

Abbildung 4.5.2–1: Neuronale Struktur eines Morphems. Die durch Lernprozesse zu verstärkenden Verbindungen sind gestrichelt dargestellt. Weitere Erläuterungen im Text.

Ein wichtiger Unterschied zwischen Silben und Morphemen ist, wie in Abbildung 4.5.2–1 im Vergleich zu Abbildung 4.5.1–3 ersichtlich, dass die einheitenbildende ODER-Zelle bei Morphemen nicht zur Ausdrucksseite, sondern

zur Inhaltsseite gehört, denn sie soll ja – für Morpheme definitionsgemäß – auch den Übergang zu den Inhalten leisten.

Wenn man beachtet, dass es in frühen Spracherwerbsphasen zu Übergeneralisierungen im morphologischen Bereich kommt (z. B. zur Regularisierung unregelmäßiger Verbformen), und wenn man zur Erklärung nicht auf die Möglichkeit von Analogiebildungen auf symbolverarbeitender Basis zurückgreifen möchte, wird ein Morphem*inventar* vorausgesetzt, nicht nur eine morphologische Segmentierung jeweils der einzelnen Lexikoneinträge. Die von Rumelhart & McClelland (1986a) vorgeschlagene konnektionistische Lösung des entsprechenden Repräsentationsproblems, die ohne ein solches Inventar auskommt, ebenso wie die modernere Variante bei Plunkett & Marchman (1991, 1993, 1996) bzw. Plunkett & Elman (1997) verbieten sich aus vielen Gründen, auf die hier nicht näher eingegangen werden soll (man vgl. allgemein Teil 2, „Grundlagen“, Abschnitte 2.2.3 und 2.2.5).

Morpheme müssen prinzipiell wie einzelne Lexikoneinträge behandelt werden. Das bedeutet, dass sie auch dann, wenn sie nicht als isolierte Wörter vorkommen, also „gebundene“ Morpheme sind, einen Startkontext bzw. ein Startsignal voraussetzen, wie das für lexikalische Ausdrucksseiten oben in Abschnitt 4.3.1 gezeigt worden ist. Man muss sich nun überlegen, wie ein solches Startsignal, gerade auch für den Fall der gebundenen Morpheme, zustande kommen kann. Bei isolierten, nach einer ausreichend langen Pause erscheinenden Wörtern kann das Startsignal aufgrund der Pause generiert werden. Das mag zunächst als die einfachste Möglichkeit erscheinen. In Abschnitt 4.3.1 ist festgestellt worden, dass ein Startsignal nur dann gegeben werden soll, wenn die bisher verarbeitete Lautkette eine lexikalisch mögliche Kette ist. Eine zweite Möglichkeit ist also, dass das Ende einer lexikalischen Ausdrucksseite ein Startsignal triggert, auch dann, wenn keine Pause vorhanden ist. Grundlage für die Entstehung des Signals ist in diesem Fall der augenblickliche Stand des lexikalischen Wissens. Auch diese Möglichkeit kann ohne besondere Probleme akzeptiert werden. Eine dritte Quelle ist weniger offenkundig, sie wird deutlich, wenn man Erscheinungen natürlicher gesprochener Sprache betrachtet. Wenn z. B. im Rahmen von Formulierungsprozessen eine Äußerung mitten in einem Wort abgebrochen wird, beginnen mögliche Fortsetzungen mit Wortanfängen, nicht mit anderen Wortfragmenten. Die Inkohärenz muss also zur Bildung eines lexikalischen Startsignals geführt haben.

Ein wichtiger zusätzlicher Gesichtspunkt ist, dass die Entstehung des Startsignals mindestens in dem zuletzt genannten Fall kein rein lexikalisches, sondern oberhalb der lexikalischen Ebene angesiedeltes Phänomen ist. Details dazu werden in Teil 5, „Syntax“ behandelt. Wenn man entsprechende

Funktionen voraussetzt, kann z. B. eine Kette wie /fiʃe/ „Fische“ in /fiʃ/ (Vollständige lexikalische Kette) plus /e/ (Pluralmorphem) zerlegt werden. Voraussetzung für die Gültigkeit der Zerlegung ist allerdings, dass beide Bestandteile tatsächlich vollständige Bedeutungsträger sind, wie es der Morphemdefinition entspricht.

Die bisherigen Überlegungen dieses Abschnitts betreffen die Funktion „fertiger“ Lexikonelemente. Wenn man an die Problematik des Silbenkonzepts zurückdenkt, ist eine entscheidende Frage jetzt die, ob die vorausgesetzten Lexikonstrukturen durch Lernprozesse zustande kommen können, vor allem, ob eine morphologische Segmentierung (im Lernprozess) bei gebundenen Morphemen erklärbar ist. Es ist viel einfacher, einen linguistischen Laien dazu zu bringen, die Form /fiʃe/ in die Silben /fi/ und /ʃe/ zu zerlegen, als in die Morpheme /fiʃ/ und /e/. Man muss sich als Linguist fragen, welche Anhaltspunkte man in einem Lernprozess für die in vielen Fällen als ungewohnt empfundene Segmentierung hat.

Man könnte versucht sein, in dem Beispiel /fiʃe/ eine Segmentierung dadurch vorzunehmen, dass man sagt, dass die Morphemgrenze dort liegt, wo die Bedeutung im Wort sich ändert, also in diesem Beispiel hinter /fiʃ/, wo die Bedeutung PLURAL sozusagen beginnt. Diese Vorstellung ist aber schon allein deshalb nicht haltbar, weil der Input bei einem Lernvorgang die Konzepte FISCH und PLURAL nicht hintereinander gereiht, sondern bestenfalls gleichzeitig enthält. Ein einzelnes Vorkommen von /fiʃe/ in einem Verstehensprozess kann deshalb nicht die Grundlage einer Segmentierung sein, was auch den allgemeinen Prinzipien einer morphologischen Analyse in der Linguistik entspricht. Ein Linguist würde mindestens das „Minimalpaar“ /fiʃ/ FISCH SINGULAR und /fiʃe/ FISCH PLURAL vergleichen.

Wenn man einmal davon absieht, dass ein getrennter Lexikoneintrag für das Morphem /e/ PLURAL entstehen sollte und man also nur eine Verteilung der Kette auf zwei Bedeutungen verlangt, was man als eine Form von Segmentierung sehen kann, und wenn man den Vergleich des Linguisten als Folge von Inputs in eine neuronale Struktur auffasst, ist das Ergebnis allerdings nur halb befriedigend. Eine Simulation mit der Architektur der Abbildung 4.5.2–2 soll das demonstrieren. Es ist nur eine einzige lexikalische Sequenz, nämlich /fiʃe/ vorhanden, zusammen mit den inhaltlichen Konzepten SINGULAR (SG), PLURAL (PL) und FISCH. Die lautliche Sequenz und die inhaltlichen Konzepte gelten als schon aufgebaut, es geht also nur noch um den Aufbau der entsprechenden Verbindungen zwischen den ausdrucksseitigen und den inhaltsseitigen Elementen. Die einzelnen lautlichen Segmente (genauer: die zugehörigen sequenzenbildenden Zellen) haben vor dem Lernprozess schwache Verbindungen mit allen inhaltlichen Konzepten

(den zugehörigen ODER-Zellen); der morphologische Lernprozess bezieht sich auf diese Verbindungen. Er entspricht insofern der Idee des Minimalpaarvergleichs, als gerade *ein* Exemplar der Pluralbedeutung und *ein* Exemplar der Singularbedeutung verwendet wird.

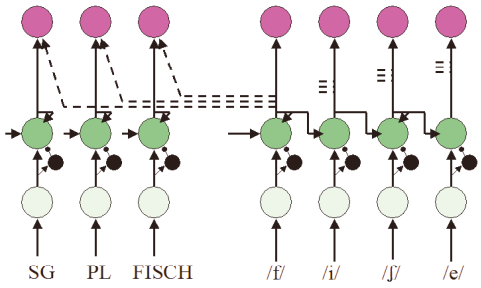

Abbildung 4.5.2–1: Einfache Architektur zur Simulation der Morphemsegmentierung von *Fisch* und *Fische*. Die gestrichelten, von sequenzenbildenden Zellen auf ODER-Zellen führenden Verbindungen, die der Übersichtlichkeit halber nur für eine der Phoneminstanzen vollständig eingetragen sind, sollten idealerweise so verstärkt werden, dass das /e/ am Wortende die Bedeutung PLURAL und der Rest die Bedeutung FISCH bekommt.

Die Simulation unterscheidet eine Lernphase, bei der in entsprechender Synchronisation (vgl. Kapitel 4.4.2) die inhaltlichen Konzepte und die beiden Ausdrucksseiten (zunächst der Plural, dann der Singular) angeboten werden, und eine Kannphase, in der die Kette /fiʃe/ allein den Input bildet. In der Kannphase kann man beobachten, dass die einzelnen Phoneme von /fiʃ/ jeweils alle vorhandenen Bedeutungskomponenten aktivieren, also SINGULAR, PLURAL und FISCH, das abschließende Phonem /e/ aber nur PLURAL und FISCH. Es ist also eine gewisse Segmentierung eingetreten, die allerdings banal ist, wenn man bedenkt, dass das schließende /e/ im Input nie zusammen mit SINGULAR vorkommt, und deshalb eine entsprechende Verbindung auch nicht aufgebaut werden kann.

Simulation:
Einfache Simulation zur morphologischen Segmentierung.
Man beachte, dass sich die Kreise, die die Verbindungsgewichte symbolisieren, teilweise überlagern.
Die Option „Simulation bis Stop“ führt an den Beginn der Kannphase.
Weitere Erläuterungen im Text.

Unerwünscht ist natürlich, dass /e/ nicht nur mit PLURAL, sondern auch mit FISCH verknüpft wird. Aber der Input in dem vereinfachten Simulationsversuch entspricht auch nicht dem Input im Rahmen des Spracherwerbs.

Der wesentlichste Unterschied besteht darin, dass im Input des Spracherwerbs das Pluralzeichen /e/ sehr viel häufiger vorkommt, als die unsegmentierte Wortform /fiʃe/. Dasselbe gilt – in weniger großem Umfang – auch für die Sequenz /fiʃ/, weil sie ja sowohl „in Isolation", als auch in der Sequenz /fiʃe/ vorkommt. Diese statistische Verteilung kann, wie schon in Kochendörfer (2002: 162 ff.) gezeigt worden ist, als Grundlage einer morphologischen Segmentierung dienen. Ergänzend zu dieser früheren Version ist jetzt aber zu beachten, dass nicht einfach eine entsprechend häufige Aktivierung einer Bedeutung, repräsentiert in einer einzelnen Instanz, zusammen mit einem ausdrucksseitigen Input ein entsprechendes Ergebnis zu liefern hat, sondern dass entsprechend den in Kapitel 4.4 entwickelten Modellvorstellungen, also bei einer Bedeutungsrepräsentation durch mehrere Instanzen, *dieselben Instanzen* der Bedeutungsrepräsentation *wiederholt* an einem Lernprozess beteiligt sein müssen, um einen Lerneffekt zu erzielen.

Was sich unter solchen natürlicheren Bedingungen ergibt, zeigt die im Folgenden beschriebene Simulation mit den Inputsequenzen /fiʃ/ „Fisch", /ʃif/ „Schiff", /tiʃ/ „Tisch", /fiʃe/ „Fische", /ʃife/ „Schiffe", /tiʃe/ „Tische".

Die verwendete Architektur entspricht prinzipiell der oben in Abbildung 4.4.2–5 wiedergegebenen, mit Trennung von Bereichen für Ausdrücke und Inhalte. Die Eingabeeinheiten am linken Rand sind so definiert, wie in Abbildung 4.5.2–2 angegeben. Es wird darauf verzichtet, das Zustandekommen der inhaltlichen Konzepte SG=SINGULAR, PL=PLURAL, TISCH, FISCH und SCHIFF aus Komponenten zu simulieren, wohl aber wird die Instanzenbildung für diese Konzepte gezeigt. Dasselbe gilt für die phonologischen Einheiten /t/, /i/, /ʃ/, /e/ und /f/. Um einen einigermaßen erträglichen Zeitaufwand für die Simulation sicherzustellen, sind die Lernparameter der instanzenbildenden und sequenzenbildenden Zellen so eingestellt, dass eine einmalige Eingabe ausreicht, um die zu lernenden Strukturen aufzubauen. Abweichend davon entstehen funktionsfähige Verbindungen auf den ODER-Zellen, soweit sie nicht fest sind, in zwei Schritten. Dieser letztere Parameter bestimmt, wieviele Wiederholungen einer Ausdrucks-Inhalts-Zuordnung erforderlich sind, um eine entsprechende Verbindung zu erreichen. Zwei Schritte sind für diesen Vorgang minimal und mit Sicherheit unrealistisch wenig. Auch diese Minimalanforderung hat den Sinn, die Simulation in erträglicher Zeit möglich zu machen. Die in Kochendörfer (2002: 165) entwickelte Regel, wonach sich die Lerneffekte aus verschiedenen Eingängen auf ODER-Zellen innerhalb eines Zeitfensters nicht summieren, sondern ein für das Zeitfenster typischer Gewinn sich auf die einzelnen Eingänge verteilt, wird als selbstverständlich vorausgesetzt und ist Bestandteil aller betroffenen Simulationen im Projekt „Kortikale Linguistik".

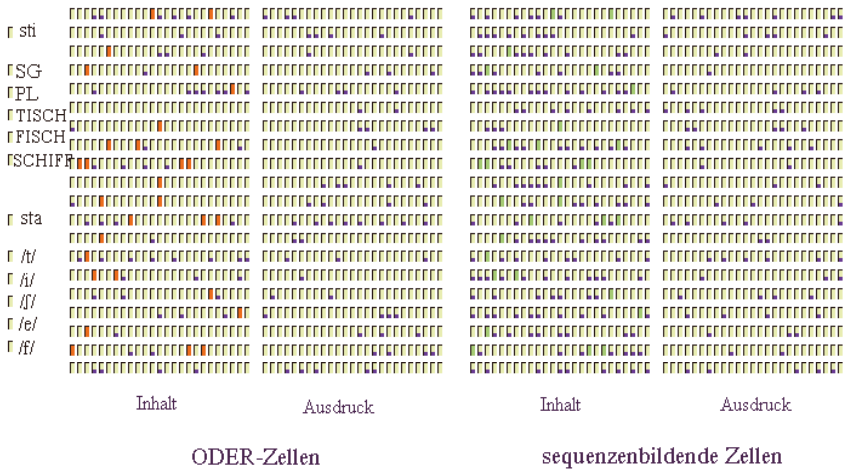

Abbildung 4.5.2–2: Architektur zur Simulation eines morphologischen Lernprozesses mit morphologischer Segmentierung. Die Abbildung gibt den Bildschirmzustand im Zeittakt 875 wieder, am linken Rand sind die Definitionen der Input-Einheiten hinzugefügt.

Der Input enthält zunächst die inhaltlichen Konzepte zur Bildung entsprechender Instanzen. Dann folgen die Singularformen /fiʃ/, /ʃif/ und /tiʃ/, in dieser Reihenfolge. Die Pluralformen schließen sich an. Wie in der vorangegangenen Simulation sind Inhalte und Ausdrücke ausreichend synchronisiert. Der Input der Wortformen geschieht mit ausreichenden Abständen, um zu verhindern, dass sich aufgrund der vereinfachenden Lernparameter Ketten über die Wortgrenzen hinaus bilden. Damit enthält der Input jeweils nur einfach die Ausdrucksseiten /fiʃ/, /ʃif/ und /tiʃ/ zusammen mit den Inhalten FISCH+SINGULAR, SCHIFF+SINGULAR und TISCH+SINGULAR, sowie die Ausdrucksseiten /fiʃe/, /ʃife/ und /tiʃe/, zusammen mit den Inhalten FISCH+PLURAL, SCHIFF+PLURAL und TISCH+PLURAL (+ bedeutet „gleichzeitig mit“); jeweils doppelt die Ausdrucksseiten /fiʃ/, ʃif/ und /tiʃ/, zusammen mit den Inhalten FISCH, SCHIFF und TISCH; das /e/ für PLURAL ist dreifach vorhanden. Da die für den Lexikonstart des morphologischen Elements /e/ erforderlichen Signale, wie oben erläutert, unter Einschaltung außerlexikalischer Prozesse erzeugt werden müssen, werden diese Startsignale an der richtigen Stelle durch einen entsprechenden externen Input eingeblendet. Auf die eigentliche Lernphase mit diesen Inputelementen folgen einige Tests, die unten noch genauer analysiert werden.

Simulation:
Morphologischer Lernprozess mit Segmentierung.

Mit der Option „Simulation bis Stop“ können die folgenden für den Ablauf interessanten Punkte direkt erreicht werden:

Zeittakt 800: Beginn der Eingabe der Singularformen;
Zeittakt 2000: Beginn der Eingabe der Pluralformen;
Zeittakt 3300: Test mit /e/ allein, nach Neustart;
Zeittakt 3500: Test mit der Eingabe /fiʃe/, extern erzeugter Neustartimpuls vor /e/;
Zeittakt 3900: Test mit /fiʃe/, ohne extern erzeugten Neustartimpuls vor /e/;

Weitere Erläuterungen im Text.

Die folgende Tabelle stellt einige exemplarische Beobachtungen während des Simulationsverlaufs zusammen.

Phase 1: Instanzen inhaltlicher Konzepte werden generiert	
Zeittakt	Kommentar
0 ff.	Ein relativ umfangreicher Startkontext für Inhalte wird erzeugt.
80	22 ODER-Zellen für SINGULAR feuern. Der inhaltliche Kontext wird erweitert bzw. neu gesetzt, aber in dieser erweiterten Form nicht genutzt, da der Input nur einmal und mit entsprechendem zeitlichem Abstand erfolgt.
220	12 ODER-Zellen mit der Bedeutung PLURAL feuern.
360	15 ODER-Zellen mit der Bedeutung TISCH feuern.
500	9 ODER-Zellen mit der Bedeutung FISCH feuern.
630	8 ODER-Zellen mit der Bedeutung SCHIFF feuern.

Phase 2: Lernen der Singularformen	
Zeittakt	Kommentar
800 ff.	Startkontexte sowohl für den inhaltlichen als auch für den ausdrucksseitigen Bereich werden gesetzt.
875	30 ODER-Zellen für die Inhalte FISCH und SINGULAR feuern.
877	12 Instanzen für /f/ (von /fiʃ/) feuern.
879	Es sind erhöhte Pegel bei einigen der ODER-Zellen der Inhaltsseite entstanden, die in Zeittakt 875 gefeuert haben; keine dieser ODER-Zellen feuert anschließend.

	Analoges gilt für die gesamte Phase der Eingabe der Singulare für jedes Eingabephonem, mit Ausnahme des /f/ von /fiʃ/ .
1417	Das in Zeittakt 1410 eingegebene /f/ von /ʃif/ bringt keine ausdrucksseitige Sequenzeinheit zum Feuern, damit entfallen auch die entsprechenden anschließenden Lerneffekte.

Phase 3: Lernen der Pluralformen	
Zeittakt	Kommentar
2000 ff.	Die Phoneme der Wortstämme bilden jetzt funktionsfähige Verbindungen mit den entsprechenden ODER-Zellen der Inhaltsrepräsentationen aus.
2080	Eine mit dem Phonem /f/ von /fiʃe/ verknüpfte Zelle im Feld der inhaltsseitigen ODER-Zellen feuert. Ein analoger Vorgang ergibt sich in der Folge für jedes im Singular vollständig bearbeitete Phonem (also mit Ausnahme des /f/ von /ʃife/). Das heißt, einige der inhaltsseitigen ODER-Zellen feuern jetzt in kurzem Abstand zweimal.
2289	Das Plural-/e/ von /fiʃe/hat zu einer unterschwelligen Verstärkung einer Verbindung (eines Verbindungsbündels) mit drei inhaltsseitigen ODER-Zellen geführt, was an einem erhöhten Potenzialpegel dieser Zellen sichtbar wird. Zur weiteren Entwicklung des Pluralzeichens in der Lernphase vgl. die Zusammenstellung unten im Text.

Phase 4: Tests	
Zeittakt	Kommentar
3300 ff.	/e/ ist als eigener Eintrag im Lexikon etabliert und kann als Pluralzeichen abgerufen werden.
3380	Zwei für die Repräsentation von PLURAL zuständige inhaltsseitige ODER-Zellen feuern.
3500 ff.	Eingabe von /fiʃe/ ohne Synchronisation mit den entsprechenden Bedeutungen, aber mit externem Neustart des Lexikons vor /e/. Die Zustände auf der Inhaltsseite, die durch den jeweiligen phonologischen Input erreicht werden, sind in Abbildung 4.5.2–3 dokumentiert.
3900 ff.	Versuch mit einer Eingabe von /fiʃe/ ohne Neustart des Lexikons vor /e/.
4187	Der Input von /e/ führt zum Feuern sequenzenbildender Instanzen, ist also korrekt an die Sequenz /fiʃ/ angeschlossen.
4189	Eine überschwellige Reaktion im Feld der inhaltlichen ODER-Zellen bleibt aus.

In Abbildung 4.5.2–3 ist das Testergebnis aufgrund der Eingabe der Ausdrucksseite von /fiʃe/ (Zeittakt 3500 ff. der Simulation) zusammenfassend wiedergegeben.

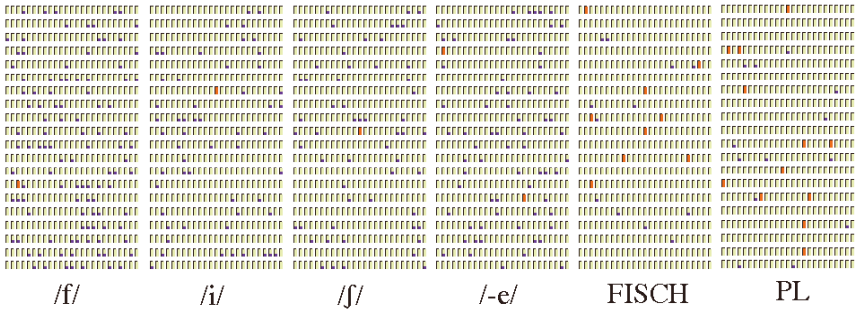

Abbildung 4.5.2–3: Aktivierte ODER-Zellen der Inhaltsseiten bei der Testeingabe von /fiʃe/ ab Zeittakt 3500. Die den einzelnen Inputphonemen entsprechenden Zellfelder sind nebeneinandergestellt. Die letzten beiden Felder geben die Aktivation der inhaltsseitigen ODER-Zellen bei der Verankerung der Inhalte FISCH und PLURAL zu Anfang der Simulation wieder. Weitere Erläuterungen im Text.

Man kann zunächst oberflächlich feststellen, dass tatsächlich die gewünschte Segmentierung geleistet worden ist, in der Weise, dass die Phoneme der Sequenz /fiʃ/ Instanzen – und zwar jeweils verschiedene Instanzen – der Bedeutung FISCH aktivieren. Dass gleichzeitig auch SINGULAR und PLURAL aktiviert werden, ist nicht zu erwarten, da /fiʃ/ zusammen mit SINGULAR und zusammen mit PLURAL jeweils nur *einmal* im Lernmaterial auftritt, und die entsprechenden Verbindungen damit nicht ausreichend verstärkt worden sind.

Das /e/ ist ausschließlich mit Instanzen der Pluralbedeutung verknüpft. Dass überhaupt eine Aktivierung mit PLURAL-Instanzen erfolgt, ist bemerkenswert, denn das /e/ ist nicht im Zusammenhang mit der Eingabe von /fiʃe/ als Träger der Pluralbedeutung definiert worden, sondern in Zusammenhang mit der Eingabe von /ʃife/ und /tiʃe/. Das ergibt sich, wenn man die schrittweise Steigerung der Summe der Effektivität der Verbindungen von /e/-Instanzen (genauer: der entsprechenden ausdrucksseitigen sequenzenbildenden Zellen) auf ODER-Zellen der Inhaltsseite genauer verfolgt:

Um den Vergleich mit der Bildschirmdarstellung zu ermöglichen, ist die Position der betroffenen ODER-Zellen (Zeile von oben gezählt/Spalte von links gezählt) angegeben.

Verstärkte Verbindungen aufgrund der Eingabe des /e/ von /fiʃe/, in Zeittakt 2289: 9/16, 12/9, 13/16.
Zusätzlich verstärkte Verbindungen aufgrund der Eingabe des /e/ von /ʃife/, in Zeittakt 2689: 4/2, 6/13, 7/18, 10/17, 15/17, 16/3.
Zusätzlich verstärkte Verbindungen aufgrund der Eingabe des /e/ von /tiʃe/ in Zeittakt 3089: 2/10, 2/17, 4/2, 7/9, 14/18, 15/8, 15/17, 16/6, 18/1, 18/6, 20/24.

Nur die Verbindungen mit den Zellen 4/2 und 15/17 sind zweimal verstärkt worden. Diese beiden Zellen feuern schon in Zeittakt 3089 ausgelöst durch die Eingabe des /e/ von /tiʃe/ und reagieren entsprechend überschwellig im Test mit /e/ ab Zeittakt 3300. Es fällt auf, dass die meisten verstärkten Verbindungen Zellen aktivieren, die nicht schon zu Anfang als Pluralrepräsentationen ausgewählt worden sind. Es handelt sich dabei um Pluralrepräsentationen, die durch den speziellen inhaltlichen Kontext möglich geworden sind. Da diese spezielle Kontextbindung die Wiederholung in der Simulation ausschließt, lässt sich das Phänomen dadurch erklären. Nur die beiden zweimal verstärkten (und zusätzlich die nur einmal verstärkte Zelle 15/8) sind auch schon zu Anfang mit der Bedeutung PLURAL verknüpft. (Man beachte, dass die Effekte von verstärkten, aber nicht feuernden Verbindungen in Abbildung 4.5.2–3 nicht sichtbar werden, da sie in dem dargestellten Zeittakt schon so weit abgeschwächt sind, dass sie nicht mehr von ganz schwachen Effekten unterschieden werden.)

Wie der Test in der Simulation ab Zeittakt 3900 ergänzend zeigt, werden die Verbindungen, die durch das sequenziell, ohne Neustart des Lexikons angeschlossene /e/ von /fiʃe/ aufgebaut worden sind, zwar benutzt, führen aber nicht zum Feuern von Instanzen der Bedeutung PLURAL.

Die Simulation zeigt, dass Morpheme, im Unterschied zu Silben, tatsächlich lernbar sind. Der Grund für die Differenz ist in der Tatsache zu sehen, dass Morpheme Bedeutungen tragen und damit als Einheiten ausgezeichnet sind. Das episodische(!) Lernen von Silben erklärt sich auf dieselbe Weise.

4.5.3 Das morphologische Lexikon

Interessante Beobachtungen ergeben sich, wenn man die im vorigen Abschnitt durchgeführte Simulation unter Beibehaltung aller Parameter fortsetzt. Im Folgenden wird dazu ein Netz verwendet, das dem Zustand entspricht, der im Zeittakt 3301 dieser Simulation erreicht ist. Der Input in dieses Netz, das also den Effekt des zurückliegenden Lernvorgangs spiegelt, besteht in der Eingabe der Singular- und Pluralformen in demselben Um-

fang und in derselben Anordnung wie oben verwendet. Es sind also die Inputs „Fisch“, „Schiff“ und „Tisch“ je einmal in der Singularform und in der Pluralform vorhanden. Tests mit dem Pluralzeichen /e/ und der Wortform /fiʃe/ schließen sich an. Da die inhaltlichen Konzepte im Ausgangsnetz schon aufgebaut sind, entfällt das Lernen dieser Konzepte und damit die erste Phase der Simulation von Abschnitt 4.5.2. Der Verlauf der Aktivationen auf der Inhaltsseite während der Eingabe von /fiʃe/ ist (entsprechend der Abbildung 4.5.2–3) in Abbildung 4.5.3–1 wiedergegeben.

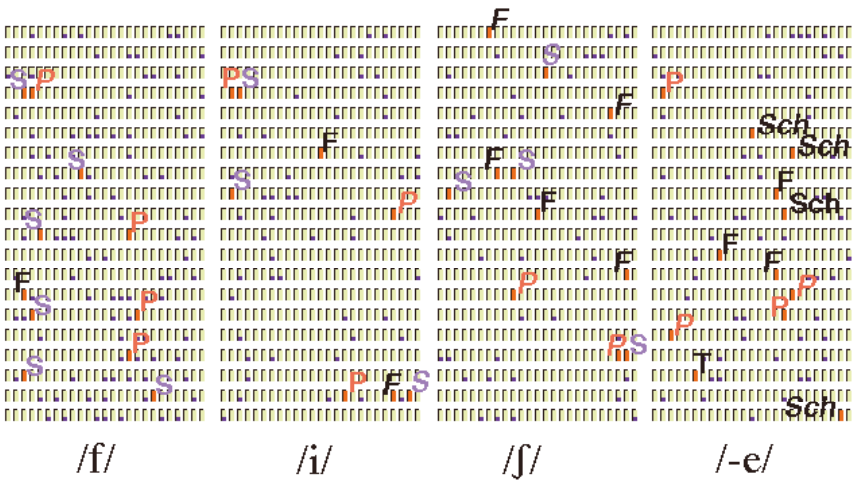

Abbildung 4.5.3–1: Aktivierte ODER-Zellen der Inhaltsseiten bei der Testeingabe von /fiʃe/ nach einer zweiten Lernrunde. Die Anordnung entspricht der Abbildung 4.5.2–3. Es sind zusätzlich die Bedeutungen der feuernden Zellen eingetragen: F = FISCH, Sch = SCHIFF, T = TISCH, S = SINGULAR, P = PLURAL. Instanzen inhaltlicher Konzepte, die vor der zweiten Lernrunde noch nicht verfügbar waren, sind kursiv ausgezeichnet.

Simulation:

Morphologischer Lernprozess, Fortsetzung.

Mit der Option „Simulation bis Stop“ können die folgenden für den Ablauf interessanten Punkte direkt erreicht werden:

Zeittakt 1200: Beginn der Eingabe der Pluralformen;
Zeittakt 2500: Test mit /e/ allein, nach Neustart;
Zeittakt 2800: Test mit der Eingabe /fiʃe/, extern erzeugter Neustartimpuls vor /e/;
Zeittakt 3300: Test mit /fiʃe/, ohne extern erzeugten Neustartimpuls vor /e/.

Das Bild ist schon rein optisch wesentlich komplexer als das mit der Wiedergabe des Ergebnisses der ersten Lernrunde oben in Abbildung 4.5.2–3.

Zur Erklärung muss man etwas weiter ausholen und sich die Parameter vergegenwärtigen, die für die Simulationen hier gegolten haben: Der Lernschritt an den verstärkbaren Verbindungen, die zu den ODER-Zellen führen, ist so eingestellt, dass der verwendete Input in der ersten Lernrunde gerade noch zu einer Segmentierung führt, das heißt, ein zweimaliges Zusammentreffen des Ausdrucks mit einem Inhalt genügt zum Aufbau einer funktionsfähigen Verbindung zwischen Ausdruck und Inhalt. Wenn man den Lernschritt so verändert, dass schon ein einfaches Zusammentreffen genügt, um eine solche Verbindung herzustellen, findet keine Segmentierung statt. Wenn man den Lernschritt so einstellt, dass eine dreifache Wiederholung erforderlich ist, hat das /e/ noch eine Chance, mit PLURAL verknüpft zu werden, aber nicht mit FISCH, SCHIFF oder TISCH. Bei Erhöhung der erforderlichen Wiederholrate wird die Differenzierung immer deutlicher, vorausgesetzt, es ist eine entsprechende Differenz der Vorkommenshäufigkeit zwischen der Pluralendung einerseits und jedem einzelnen Stamm andererseits vorhanden.

Wenn man annimmt, dass die Stämme immer und gleich häufig entweder mit SINGULAR oder PLURAL zusammen erscheinen, kann nicht verhindert werden, dass sie eine SINGULAR-PLURAL-Mehrdeutigkeit entwickeln, es sei denn, die Stämme sind im Input verschieden, das heißt, sie erscheinen in verschiedenen Formen. Wenn die Stämme im Input verschieden sind, können sie auch zusätzlich die Information des Numerus tragen, also *entweder* SINGULAR *oder* Plural, gleichgültig, ob ein spezialisiertes Numerusmorphem sequenziell folgt oder nicht. Wenn man für das Entstehen eines Lerneffekts generell eine bestimmte Häufigkeit verlangt, sind Wörter, die häufig im Input erscheinen, bezüglich dieser Erscheinung bevorzugt, da die Zahl der Tokens für das einzelne Allomorph gegenüber dem Morphem z. B. für zwei Allomorphe entsprechend halbiert wird.

Häufig im Input erscheinende Wörter können schließlich auch Vollformen ausbilden. Vollformen bestehen in vollständigen ausdrucksseitigen Sequenzen, sie entstehen nicht durch Zusammenrücken von Morphemen. Dasselbe gilt für Wortbildungsprodukte. Die Annahme von Strukturen, wie in 4.5.3–2 dargestellt, stößt auf das schon in Kochendörfer (2002: 168) festgestellte Problem, dass eine in dieser Weise gebildete Kette von Morphemen auch schon aufgrund eines bruchstückhaften Inputs vollgültig aktiviert werden würde, und zwar in der Weise, dass von den Ausdrucksseiten der einzelnen Morpheme der Kette jeweils nur das Anfangsphonem oder ein aussprechbares Bruchstück vorhanden sein müsste, in dem Beispiel der Abbildung

4.5.3–2 also z. B. /titu:/. Man beachte, dass man keine Möglichkeit hätte, diese Gebilde als nicht-lexikalisch zu identifizieren. Jedes einzelne Phonem von /tiʃ/ würde die Zellen TISCH und TISCHTUCH überschwellig erregen und dasselbe gilt entsprechend für /tu:x/, TUCH und TISCHTUCH. Die Sequenzkontrolle bleibt wirkungslos.

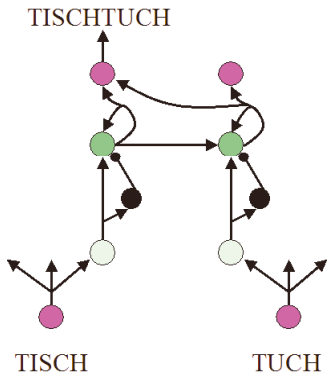

Abbildung 4.5.3–2: Schema zur Diskussion der Möglichkeit von Morphemketten im Lexikon.

Die unerwünschten Phänomene können nur dadurch vermieden werden, dass verlangt wird, dass das Ende des ersten Bestandteils der Morphemsequenz abgewartet werden muss, ehe der zweite Bestandteil „gestartet“ wird. Bedeutungen können aber kein definiertes Ende haben. Eine durch entsprechende neuronale Strukturen gewährleistete Ende-Information gehört, wie in Kochendörfer (2002) gezeigt worden ist (man vgl. auch unten 4.6.1 und Teil 5, „Syntax“), zum syntaktischen Bereich. Die Syntax muss aber generell verantwortlich sein für den Start lexikalischer Sequenzen. Der Start von zweiten Bestandteilen von Kompositia, wie in dem hier verwendeten Beispiel, muss mit dem allgemeinen Lexikonstart identisch sein. Die Konsequenz ist, dass Morphemsequenzen der beschriebenen Art für das mentale Lexikon nicht angenommen werden dürfen.

Das Lexikon einer Sprache enthält definitionsgemäß Einheiten, die ihrerseits in dem Sinne komplex sind, als sie prinzipiell aus einer zeitlichen Sequenz von Einheiten (Lauten) bestehen, die selbst nicht mehr in der gleichen Weise segmentierbar sind. Nach den in den vorangegangenen Abschnitten 4.5.1 und 4.5.2 angestellten Überlegungen wird die Einheit einer lexikalischen Sequenz durch die Bedeutung hergestellt, die eine solche Sequenz hat. Zusätzlich gibt es Beziehungen zur Syntax. Das mentale Lexikon ist in dem Sinne morphologisch, dass es ausschließlich aus elementaren Morphemen besteht.

Eine morphologische Zerlegung von Wörtern (das heißt, die Verteilung eines Worts auf verschiedene lexikalische Einheiten, also Einheiten, die einen eigenen Lexikonstart voraussetzen) ist möglich, das heißt aber nicht, dass sie unter allen Umständen auch passiert. Es besteht eine wichtige Abhängigkeit von den statistischen Eigenschaften des Inputs. Man beachte, dass „Input“ auch durch inneres Sprechen während eines Lernprozesses erzeugt werden kann. Die folgende Zusammenstellung bezieht sich auf das Deutsche und strukturell ähnliche Sprachen.

- Wenn die Häufigkeit des Auftretens einer Form im Input groß genug ist, wird sie, selbst dann, wenn eine morphologische Zerlegung möglich oder bereits erfolgt ist, als Vollform gespeichert.

- Wörter mit Eindringen von Flexionsinformationen in den Stamm sollten ebenfalls zu den im Input häufigeren Wörtern gehören. Dabei muss wahrscheinlich nicht unterschieden werden zwischen Wörtern, die ein Flexionsmorphem besitzen und solchen, bei denen der Stamm allein der Träger der entsprechenden Information ist.

- Seltene Wörter erscheinen prinzipiell segmentiert. Das kann auch zur Folge haben, dass ein selten gebrauchtes oder ein neues Wort mit einer beliebigen oder einer als „Default“ gängigen morphologischen Ergänzung erscheint. Die Abhängigkeit eines Flexionselements von einer bestimmten lautlichen Eigenschaft des Stamms ist durch sprachliche Tradition bestimmt und muss nicht durch eine neuronal besonders ausgestattete Regelhaftigkeit erklärt werden.

- Die Verankerung von Allomorphen, das heißt von verschiedenen Ausdrücken für genau dieselbe Bedeutung, stellt, entsprechende Inputfrequenz vorausgesetzt, kein irgend besonderes bzw. bemerkenswertes Problem dar.

Es ist wichtig, festzuhalten, dass die unterschiedlichen Repräsentationsformen nicht zu unterschiedlichen Zugriffsprozeduren oder „Routen“ führen. Der Lexikonzugriff ist völlig einheitlich. Er beginnt mit dem syntaktisch ausgelösten Lexikonstart, der die Voraussetzung für den Beginn des geordneten Abgleichs mit dem ausdrucksseitigen Input schafft. Das Ende einer lexikalischen Kette, das eine syntaktische Information darstellt, löst einen neuen Lexikonstart aus, gleichgültig, ob die zu Ende gegangene oder die neu zu startende Kette ein freies oder gebundenes Morphem darstellt. Erforderliche Steuerungen in diesem Bereich sind nicht mehr Sache des Lexikons. Es

ist vor allem auch nicht möglich, auf einer prälexikalischen(!) „Zugriffsebene“ mit einer morphologischen Analyse zu rechnen, die dann zu verschiedenen Lexikonzugriffen führen würde (für einen Überblick über entsprechende Vorstellungen vgl. Sonnenstuhl-Henning, 2003).

4.6 Schnittstellen und Konsequenzen

4.6.1 Schnittstellen

Phonologie

Keine der in den vorangegangenen Kapiteln beschriebenen Funktionen kann gewährleistet werden, wenn man nicht akzeptiert, dass der Output der Phonologie im Verstehensprozess pro wahrgenommenem Phonem in einem einzelnen Impuls besteht, der auf jeder Zelle ausgelöst wird, die als Repräsentation des Phonems dient („Einzelimpulskodierung“). Lange und gelängte Phoneme werden entsprechend durch mehrere Impulse darsgestellt. Die erforderlichen Lernprozesse setzen, von diesem grundsätzlichen Kodierungsprinzip abgesehen, außerdem voraus, dass Phoneme im Bereich des mentalen Lexikons sehr viele gleichberechtigte Repräsentationen (Instanzen) ausbilden.

Wenn man davon ausgeht, dass an der Sinnesperipherie Frequenzkodierung herrscht, werden für die Phonologie entsprechende Umkodierungsprozesse vorausgesetzt. Analoges wird nun auch für die lexikalische Produktion vorausgesetzt. Sie liefert ebenfalls einzelimpulskodierte Information, die, unter der Voraussetzung, dass auch für die Ansteuerung von Muskeln Frequenzkodierung erforderlich ist, eine entsprechende Umwandlung auf dem Weg zur Artikulation erfahren muss.

Das lexikalische Geschehen verläuft im ungestörten Fall zeitlich streng synchron mit den Vorgängen an der Sinnesperipherie und der Motorik. Das bedeutet, dass die Schnittstellen nicht durch „Buffers“ gebildet werden müssen. Man vergleiche zu diesem Punkt den Abschnitt 4.6.5 unten. Es kann auch weder für die Produktion noch für die Perzeption eine Zwischenschicht von Teilketten (z. B. Silben) angenommen werden.

Die phonologische Segmentierung im Verstehensprozess ist prälexikalisch (siehe oben), lexikalisch nur insofern, als Mehrdeutigkeiten bei einzelnen

Phonemen zu parallel repräsentierten verschiedenen lexikalischen Inputs führen und solche Mehrdeutigkeiten dann durch den lexikalischen Prozess entschieden werden.

Prosodie

Eine Schicht der prosodischen Kontur eines Satzes oder eines Textes wird durch lexikalische Eigenschaften gebildet. Lexikalisch ist z. B. der Wortakzent. Es muss überlegt werden, wie solche lexikalischen Eigenschaften repräsentiert und realisiert zu denken sind. Mindestens zwei grundsätzlich verschiedene Möglichkeiten müssen dabei in Betracht gezogen werden:

(1) Die Prosodie kann der phonologischen Struktur inhärent sein, in dem Sinne, dass die lautlichen Einheiten je für sich ihren prosodischen Status implizieren. Das bedeutet z. B., dass das Wort „Kanne“ ein betontes /a/ enthält und dieses Phonem (diese Phoneminstanz) nicht identisch ist mit einem unbetonten /a/.

(2) Die Prosodie kann eine eigene Struktur bilden, die von der phonologischen Struktur eines Worts verschieden ist.

Unter der Voraussetzung, dass die lautliche Ebene in der Verstehensarchitektur phonologische und prosodische Informationen getrennt liefert (und nur dann!), ist es durchaus denkbar, dass Lernprozesse auch separate prosodische Muster aufbauen (Variante 2), unter derselben Voraussetzung ist es aber auch möglich, dass im lexikalischen Bereich Instanzen gebildet werden, die prosodische und phonologische Informationen verbinden (Variante 1). Eine Entscheidung ergibt sich erst, wenn man beachtet, dass die Aktivierung der in der Variante (2) entstandenen Ketten für ein Wort bei der Produktion dieses Worts streng synchron ablaufen müsste. Dabei dürften längbare Phoneme eine besondere Schwierigkeit darstellen. Es ist unklar, wie dergleichen prinzipiell zu erreichen wäre.

Syntaktische Informationen, Schnittstelle zur Syntax

Syntaktische Informationen sind mit lexikalischen Ausdrucksseiten verknüpft, aber weniger spezifisch für den einzelnen lexikalischen Ausdruck als es die (semantischen) Inhaltsseiten sind. Sie sind insofern syntaktisch, als sie den Einbau des lexikalischen Elements in eine syntaktische Struktur steuern. Es ist nicht impliziert, dass eine solche syntaktische Struktur als Datenelement aufgebaut und in einem Kurzzeitspeicher abgelegt wird. Wie in

Kochendörfer (2002) dargestellt (vgl. auch Teil 5, „Syntax“), rechnet auch das (im Spracherwerb zunächst prosodisch bestimmte) Ende einer lexikalischen Kette zu den syntaktischen Informationen. Dieses letztere Moment zwingt dazu, Bahnen, die von der lexikalischen Ausdrucksseite zur Syntax führen, zu trennen von den Bahnen, die den semantischen Gehalt anschließen (Abbildung 4.6.1–1).

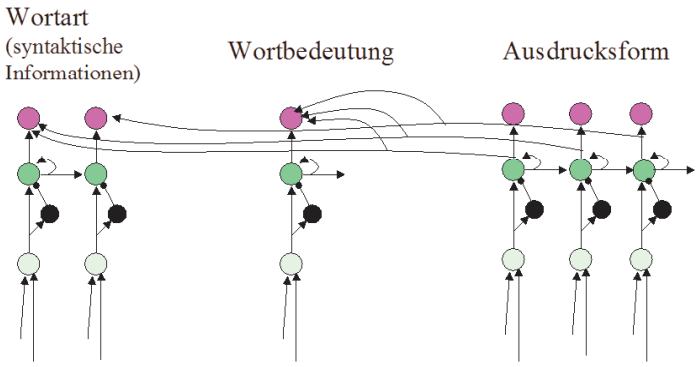

Abbildung 4.6.1–1: Lexikalische Anbindung von syntaktischen Informationen.

Die erforderlichen Lernprozesse zur Verstärkung der entsprechenden Verbindungen zur Syntax in Perzeptionsrichtung entsprechen denen zur Anbindung der lexikalischen Inhaltsseiten. Die Funktion der syntaktischen Informationen in der Produktion muss von dem allgemeinen Problem der syntaktischen Produktion her diskutiert werden und ist Teil 5, „Syntax“ vorbehalten.

Schnittstelle zum episodischen Gedächtnis

Das Charakteristikum des episodischen Langzeitgedächtnisses ist nicht eine besondere Struktur, sondern vor allem eine besonders rasche Form des Einspeicherns (vgl. Kochendörfer, 1999). Jedes Konzept kann (auch zusätzlich) episodisch verankert werden. Da dabei immer auch ein situativer Kontext wirksam sein kann, erklärt sich relativ einfach der gängige Hinweis, dass die Inhalte des episodischen Gedächtnisses im Unterschied zu anderen Gedächtnisformen mit einem Raum- und Zeitbezug versehen sind.

Sprachlich vermittelte Inhalte können in einem Verstehensprozess problemlos zum Gegenstand episodischer Spuren werden, ganz abgesehen davon, dass der Aufbau von lexikalischen Inhaltsseiten episodische Speicherprozes-

se einschließen wird. In keinem Fall darf man sich eine Art Informationstransport vorstellen, der lexikalisch verankerte Inhalte in einen episodischen Langzeitspeicher übertragen würde. Es kann immer nur darum gehen, dass vorhandene noch nicht funktionsfähige Verbindungen zu einer Erweiterung des vorhandenen Repräsentationsbestands genutzt werden.

Lesen

Lesen kann nicht ausschließlich darin bestehen, dass Grapheme oder Graphemkombinationen nach Regeln in Phoneme umgesetzt werden und damit eine phonologische Ausdrucksseite gewonnen wird, die dann der Schlüssel zu den lexikalischen Inhalten wäre. Man denke an Sprachen mit Schrifttraditionen, die einen größeren Abstand zwischen geschriebenen und gesprochenen Ausdrucksseiten zulassen oder an die Tatsache, dass man auch orthographisch entstellte Wörter (Druckfehler!) in vielen Fällen problemlos und oft ohne eine Störung zu bemerken lesen kann. Es muss also die Möglichkeit bestehen, dass optische Ausdrucksseiten direkt mit Inhalten verknüpft werden können. Soweit solche optischen Ausdrucksseiten sequenziell organisiert sind, müssen die einzelnen Sequenzelemente nicht den lautlichen Elementen des betreffenden Worts entsprechen. Das kann wichtig sein zur Erklärung der Tatsache, dass geübte Leser schneller lesen können, als sie sprechen. Beim lauten Vorlesen kann die auf dem visuellem Weg aktivierte Inhaltsseite eines Worts die Basis für eine unauffällige Produktion sein und inneres Sprechen kann einen Lesevorgang immer begleiten. Voraussetzung ist, dass die Zuordnung von optischem Muster und Inhaltsseite in einem Lernprozess hergestellt worden ist.

Gesunde Leser können aber auch Nicht-Wörter lesen, also optische Ausdrucksseiten lautlich oder durch inneres Sprechen realisieren, die nicht durch einen Lernprozess verankert sind. Das ist nur möglich, wenn Teile des optischen Musters über Lernvorgänge direkt mit lautlichen Äquivalenten verbunden worden sind. Der Umweg über die Inhaltsseiten braucht also nicht etabliert zu sein, um einen – allerdings nicht mit Sinnverständnis verknüpften – Leseprozess zu ermöglichen. Unterschiedliche „Leserouten“ sind in der Neurolinguistik auf der Basis von Mortons Logogen-Modell (vgl. oben Abschnitt 4.1.2) ausführlich diskutiert worden. Einen Überblick gibt z. B. Graap (1998).

Die für das Lesen relevanten Lernprozesse und Funktionen sollten in das hier vorgeschlagene neuronale Modell des Lexikons problemlos integriert werden können. Die Details übersteigen wegen der zusätzlich zu klärenden optischen Wahrnehmungsvorgänge den rein sprachlichen Bereich.

4.6.2 Priming-Effekte

Priming-Effekte gehören zu den stabilsten Phänomenen in psycholinguistischen Experimenten. Unter den verschiedenen Varianten ist in unserem Zusammenhang besonders das „automatische semantisch-assoziative“ Priming mit kurzen Interstimulusintervallen von Interesse. Den Grundgedanken stellt die Abbildung 4.6.2–1 dar. Einem „Prime“ (z. B. einem auf einem Bildschirm gezeigten Wort) folgt ein „Target“ gleicher oder verschiedener Modalität. Es werden zwei Fälle unterschieden: Der Fall, dass Prime und Target semantisch oder assoziativ relationiert sind und der Fall, bei dem das nicht (oder nicht in demselben Ausmaß) gilt. Was schließlich interessiert, ist die Differenz der Reaktionszeiten bei entsprechenden Aufgabenstellungen (z. B. bei der Frage, ob es sich bei dem Target um ein Wort handelt oder nicht) in dem einen, verglichen mit dem anderen Fall. Eine häufig beobachtbare Schwierigkeit ist natürlich, geeignete und möglichst zweifelsfreie dafür geeignete Stimuli zu finden.

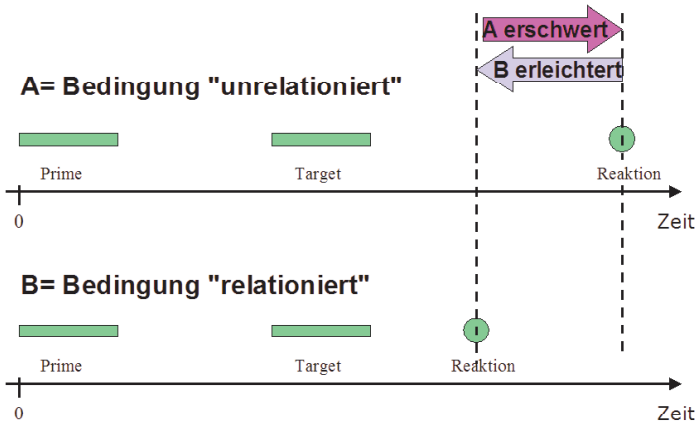

Abbildung 4.6.2–1: Schema zum Verlauf von Priming-Experimenten.

Die Differenz der Reaktionszeiten wird generell so interpretiert, dass die Bearbeitung des Targets im relationierten Fall durch den vorangegangenen Prime „erleichtert“ wird, und nicht umgekehrt, dass die Bearbeitung des Targets im unrelationierten Fall „erschwert“ ist. Den Hintergrund für diese Interpretation bildet die „Spreading-activation-Theorie“ der Semantik (z. B.Collins & Loftus, 1975). Inhaltliche Konzepte gelten als generell untereinander so verknüpft, dass ein „aktiviertes“ Konzept gradierte Aktivation an benachbarte Konzepte weitergeben kann. Wenn man einen einzelnen

Konzeptknoten betrachtet, ergeben sich Anordnung und Funktion wie in Abbildung 4.6.2–2 dargestellt.

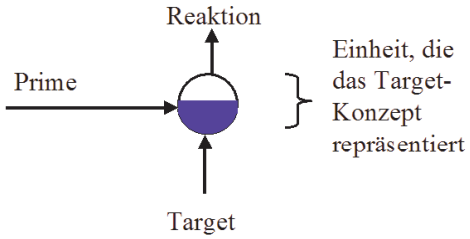

Abbildung 4.6.2–2: Elementare Architekturvorstellung zu Primingeffekten.

Ein Konzeptknoten erhält in Perzeptionsrichtung nicht nur Input von Knoten, die näher zur Sinnesperipherie liegen (über die mit Target bezeichnete Verbindung), sondern auch Input von semantisch benachbarten Knoten (über die mit Prime bezeichnete Verbindung).

Im Allgemeinen steht hinter der Spreading-activation-Idee die Vorstellung, dass semantische Konzeptknoten ein flaches Netz bilden und die Bedeutungsverwandtschaft durch die Netzverbindungen repräsentiert wird, wie in Abbildung 4.6.2–3 **A** dargestellt.

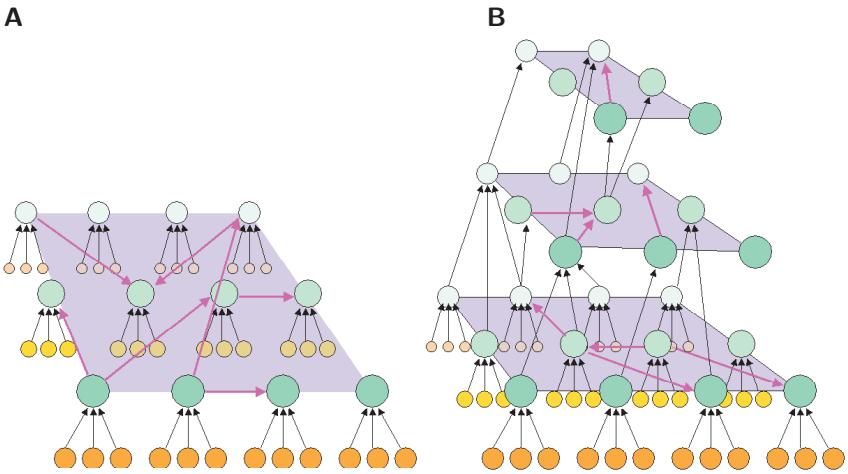

Abbildung 4.6.2–3: Annahmen zur Hierarchisierung von Konzepten.

Will man solche Vorstellungen in Einklang mit den Voraussetzungen eines biologischen Modells bringen, wird man versuchen, die Verarbeitungseinheit in Abbildung 4.6.2–2 mit einer einzelnen Zelle (im Sinne der Großmutterzellentheorie) zu identifizieren. Da die einzelne Zelle aber keine Bedeutung tragen kann, muss eine Hierarchie von Zellen dahinter stehen, die einen Kategorisierungsapparat bilden, entsprechend 4.6.2–3 **B.**

(Es versteht sich von selbst, dass Hierarchien z. B. dadurch gebildet werden, dass Axone, die den Kortex verlassen, wieder in ihn hineinführen, nicht dadurch, dass der Kortex senkrecht zur Oberfläche selbst in solche Hierarchien strukturiert ist.)

Die ein Konzept repräsentierende Zelle in einem Spreading-activation-Modell ist dann also einerseits in eine Hierarchie von Großmutterzellen eingebunden, andererseits wird sie von gleichrangigen Zellen erregt oder jedenfalls von Zellen, die zu anderen Hierarchien gehören, die dann den Priming-Effekt bewirken.

Obwohl die neuronalen Strukturen oberflächlich gesehen sehr einfach und problemlos erscheinen, lohnt es doch, eine Simulation zusammenzustellen, die das Funktionieren in der Zeit genauer beleuchtet. Die Simulationsarchitektur der Abbildung 4.6.2–4 erweitert das Modell der Abbildung 4.6.2–2 um einige Zellen, die die hierarchische Einbindung der eigentlichen „Verrechnungseinheit“ realisieren.

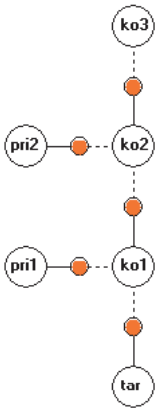

Abbildung 4.6.2–4: Simulationsarchitektur.

Die Simulationsparameter sind so gewählt, dass im unrelationierten Fall drei Impulse der Eingabezelle *tar* im Abstand von 30 Zeittakten ein überschwelliges Signal in der Zelle *ko1* bewirken. Drei Impulse dieser Zelle bewirken

wiederum ein Feuern von *ko2*. Im relationierten Fall sollte eine Folge von Impulsen von *pri1* bzw. *pri2* ausgehend einen Vorwärmeffekt auf *ko1* bzw. *ko2* erzeugen. Da nur der Zustand vor Beginn des Target-Inputs interessiert, ist in der Simulation diese „vorwärmende“ Impulsfolge auf einen einzelnen Impuls entsprechender Stärke verkürzt.

Simulationen:
Prime und Target unrelationiert.
Prime *pri1* und Target relationiert.
Prime *pri2* und Target relationiert.

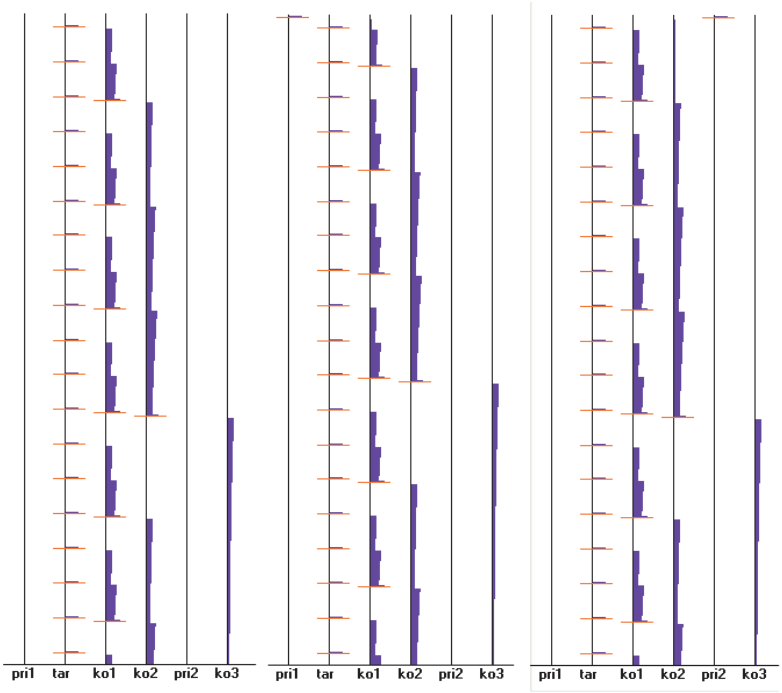

Abbildung 4.6.2–5: Simulationsergebnisse. Links: Prime und Target unrelationiert, also kein Aktionspotenzial auf *pri1* bzw. *pri2*. Mitte: Prime *pri1* und Target relationiert, also ein Aktionspotenzial auf *pri1*, die Aktionspotenziale auf *ko1* und *ko2* erscheinen früher. Rechts: Prime *pri2* und Target relationiert, das Aktionspotenzial auf *pri2* bleibt aber praktisch wirkungslos.

Die Simulationen zeigen einerseits wenigstens zu einem Teil den erwarteten „Vorwärmeffekt“, andererseits gibt es aber auch bemerkenswerte „Neben-

wirkungen“. Eine interessante Beobachtung ist, dass die Reaktionszeitunterschiede offenbar immer dem Abstand der Inputimpulse entsprechen oder ein Vielfaches davon sein müssen. Zu den Nebenwirkungen gehört, dass die Potenziale auf *ko3* immer unterschwellig bleiben, da die Abstände der Erregungen auf dieser Zelle so groß sind, dass jeweils zwischenzeitlich das erregte EPSP zu weit abgeklungen ist. Daraus könnte man die Folgerung ziehen, dass die Abnahmerate des EPSP wesentlich geringer eingestellt werden müsste. Es handelt sich aber dabei um einen durch Lernprozesse wahrscheinlich nicht veränderbaren Parameter, der also für alle Zellen eines Typs im Wesentlichen gleich sein sollte. Wenn man diesen Parameter also zugunsten von *ko3* in dem erforderlichen Ausmaß anpasst, wird auch die Verarbeitung an *ko1* und *ko2* – und hier in unerwünschter Weise – verändert.

Eine weitere wichtige eher beunruhigende Beobachtung ist, dass das Priming über *pr2* nicht zu einer Erleichterung führt. Es handelt sich wieder um ein Problem, das durch die zu rasche Abnahme des EPSP verursacht ist. Der durch den Prime ausgelöste Effekt in *ko2* ist schon zu weit abgeklungen, um noch eine nennenswerte Wirkung zu haben.

Schließlich muss auch noch beachtet werden, dass die Weiterverarbeitung eines Stimulus, ob geprimet oder nicht, dramatisch verzögert ist, verursacht durch die dem Primingeffekt zugrundeliegende Frequenzkodierung. Es gilt: Je konkreter ein Konzept ist, desto später wird es erregt. Die Verzögerung bei *ko2* beträgt über 300 Zeittakte, also mehr als das Zehnfache des Primingeffekts. Solche Zeitspannen sind für die lexikalische Verarbeitung unerträglich und machen ein Monitoring durch Überwachung einzelner Zellen unmöglich, da sich Inkohärenzen nicht mehr durch Ausbleiben von Reaktionen in einem festen Zeitfenster zeigen (man vgl. zu diesem Problem auch die Ausführungen zur inkrementellen Produktion unten in Abschnitt 4.6.5).

Man hätte sich diese ganzen Überlegungen zu den Folgen des Priming durch Spreading-Activation auch sparen können und sich zurückziehen können auf die einfache Feststellung, dass Spreading-Activation nicht möglich ist, wenn man das Prinzip der Einzelimpulskodierung nach Teil 2, „Grundlagen“, Abschnitt 2.3.2 als für den Kortex grundlegend akzeptiert. Die große Stabilität der Ergebnisse von Priming-Experimenten kann aber doch zu denken geben und so den Ausflug in eine Grundlagendiskussion rechtfertigen. Es ist außerdem ja doch so, dass die Erklärung von Priming durch Spreading-Activation regelmäßig in der einschlägigen Fachliteratur wiederholt wird, vielleicht auch einfach in Ermangelung einer Alternative. Letzteres bedeutet natürlich, dass nach der Ablehnung dieser Erklärung nun auch noch eine neue Erklärung im Rahmen eines biologisch fundierten Modells gefunden werden muss.

Der wichtigste Anhaltspunkt für eine solche neue Erklärung sind die relativ großen Reaktionszeitunterschiede, die in den Primingexperimenten festgestellt werden. Wenn Differenzen von mehr als 50ms auftreten, dann handelt es sich um Zeitspannen, die der Dauer eines Kurzvokals entsprechen könnten. Wenn man an die lexikalische Verarbeitung denkt, kann die Ursache solcher Differenzen nicht in einem längeren oder kürzeren Verlauf eines einzelnen Verarbeitungsvorgangs von der Sinnesperipherie zu einer zentralen Instanz bestehen, die das Ende des bottom-up-gerichteten Verarbeitungspfads bildet. Ein solcher Vorgang kann *insgesamt* vielleicht 20ms in Anspruch nehmen, Verarbeitungsdifferenzen in der Größenordnung der Primingeffekte können dabei nicht auftreten.

Eine weitere wichtige Überlegung betrifft den Unterschied zwischen relationiertem und unrelationiertem Stimulus: Ein Zusammenhang zwischen zwei Stimuluskonzepten kann darin bestehen, dass sie Komponenten in ihren Kategorisierungsmechanismen teilen. Das Fehlen oder Vorhandensein eines solchen Zusammenhangs kann aber für sich genommen keine nennenswerten Beschleunigungs- oder Verzögerungsphänomene verursachen. Eine zweite Form von Zusammenhängen kann über die Sequenzverbindungen zwischen den Konzepten hergestellt werden. Ein Konzept kann sequenzieller Kontext für ein zweites Konzept sein oder zwei Konzepte können denselben Kontext haben. Der unrelationierte Fall würde dann vorliegen, wenn keine dieser beiden Möglichkeiten gilt.

Jetzt ist hinzuzunehmen, dass, wie sich aus den Überlegungen oben in den Abschnitten 4.4.2 und 4.4.3 (man vgl. auch Teil 2, „Grundlagen“, Abschnitt 2.4.6) ergibt, ein passender Kontext für die inhaltliche Weiterverarbeitung lexikalischer Informationen eine wesentliche Voraussetzung ist. Wenn ein passender Kontext nicht vorhanden ist, muss er durch Kontextumschaltungsvorgänge hergestellt werden. Soche Vorgänge brauchen Zeit: Die Feststellung eines Kontextproblems geschieht dadurch, dass ein Signal nicht innerhalb eines vordefinierten Zeitfensters eine zentrale Einrichtung (als „Watchdog“ bezeichnet, man vgl. auch hier wieder den Abschnitt 4.6.5) erreicht. Das Zeitfenster muss abgewartet werden. Auch die anschließende Auslösung von kontexterweiternden Signalen und deren Wirksamwerden wird einige Millisekunden in Anspruch nehmen.

Die folgende Simulation macht weitere Bedingungen deutlich. Sie ist an die Simulation von Vorstellungsprozessen in Teil 2, „Grundlagen“, Kapitel 2.6.1, angelehnt. Dort werden die wesentlichen Funktionen erklärt: Die Feststellung der Inkohärenz des Verarbeitungsprozesses löst ein doppeltes Reparatursignal auf einem unspezifischen Verbindungsnetz aus („unspezifisches prädiktives System“, in der Simulation von der Zelle *dkt* ausgehend),

das die P-Zellen zum Feuern bringt. Gegenüber der Architektur zu Vorstellungsprozessen ist der Startkontext auf die Repräsentation des Primes erweitert und alle P-Zellen sind top-down verknüpft. Die Rückmeldungsverbindungen, die während der Perzeption unwirksam sein müssen, sind weggelassen. Die Bedingungen „relationiert“ und „unrelationiert“ sind dadurch realisiert, dass in der Bedingung „unrelationiert“ die Sequenzverbindung, die aus der Aktivierung des Primes einen Kontext für das Targetkonzept ableitet, gekappt ist.

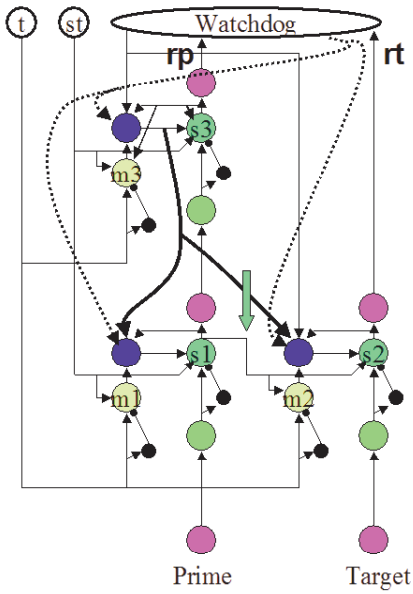

Abbildung 4.6.2–6: Simulationsarchitektur zur Demonstration von Primingeffekten durch Kontexterweiterung. Der Pfeil zeigt auf eine Sequenzverbindung, die bei der Bedingung „unrelationiert“ gekappt ist. Das „unspezifische prädiktive System“ ist gestrichelt eingetragen. Die in die Watchdog-Einheit hineinführenden Verbindungen rp und rt können zur Auslösung motorischer Reaktionen dienen. Differenzen des Aktivationszeitpunkts von rt bei den verschiedenen Testbedingungen können dann als Primingeffekte interpretiert werden.

Simulationen:
Primingeffekte durch Kontexterweiterung:
Prime und Target unrelationiert.
Prime und Target relationiert.

Bei der in der Simulation realisierten Hypothese zum Zustandekommen von Primingeffekten steht sichtlich nicht mehr der Aspekt der Beschleunigung eines Vorgangs durch einen Vorwärmeffekt im Vordergrund, sondern die Verlangsamung durch ein Inkohärenzproblem. Es ist natürlicher, von einer Verlangsamung des unrelationierten als von einer Beschleunigung des relationierten Stimulus zu sprechen. Die Verlangsamung im Einzelfall muss mindestens so groß sein wie der Abstand zweier Aktionspotenziale auf der Zelle, die das Target-Konzept repräsentiert. In der Simulation sind es 70ms. Größere oder kleinere Mittelwerte können z. B. durch mangelnde Homogenität und unterschiedlichen Schwierigkeitsgrad des Stimulusmaterials verursacht werden.

Eine Stütze für diese Erklärung ergibt sich aus Priming-Experimenten, die mit Alzheimer-Patienten durchgeführt worden sind. In diesen Experimenten zeigen die Patienten ein Phänomen, das als „Hyperpriming“ bezeichnet worden ist, das heißt, der Abstand der Reaktionszeiten ist größer als bei gesunden Probanden. Die unter der Hypothese der Spreading-Activation stehenden Erklärungen sind, wie in Nitt et al. (2005) dargestellt, wenig plausibel. Wenn man aber mit einer verlangsamten Kontextverarbeitung rechnet, passt das Verhalten in das Krankheitsbild und entspricht den Kontextverarbeitungsproblemen, die man auch sonst bei den Patienten beobachtet (vgl. Teil 9, „Sprachpathologie“).

4.6.3 Wortfrequenz-Effekt

Einige der oben in Abschnitt 4.2.6 aufgeführten experimentellen Effekte können relativ leicht geklärt werden. *Phonemrestaurationseffekte* können, soweit sie bei den Hörern unbemerkt bleiben, dadurch zustande kommen, dass die Phonologie im Störfall mehrere alternative Phoneme aktiviert und der lexikalische Abgleich dann eine Entscheidung liefert. Das Bewusstwerden des „restaurierten“ Phonems bzw. der Tatsache des Effekts setzt darauf aufbauend dann wohl einen anschließenden Vorstellungsprozess voraus. Auffällige, das heißt also nicht unbemerkt bleibende phonologische Störungen müssen grundsätzlich über Vorstellungsprozesse repariert werden. *Reizdegradierungseffekte* spiegeln ähnliche Vorgänge wie die Phonemrestaurationseffekte. *Kontexteffekte* sind schon im Zusammenhang mit Mehrdeutigkeitsphänomenen oben in Abschnitt 4.4.3 behandelt worden.

Das Zustandekommen des *Wortfrequenz-Effekts* ist dagegen nicht sofort einsichtig. Es gibt keine Möglichkeit, im Gehirn mit unterschiedlichen Schwellen für Wortdetektoren zu rechnen, wie das teilweise in der einschlägigen Literatur geschieht. Das ergibt sich schon allein aus der Ablehnung von

Frequenzkodierung in den betroffenen kortikalen Prozessen. Es ist natürlich auch nicht möglich, so etwas wie eine serielle Suche nach der Idee von Forster realisiert zu denken, es lassen sich generell keine neuronalen Lernprozesse angeben, die eine nach Häufigkeit geordnete Liste lexikalischer Ausdrucksseiten liefern würden.

Die Häufigkeit eines Worts im Input (spezifisch jeweils für eine Person) kann zur Konsequenz haben, dass mehr ausdrucksseitige Instanzen dieses Worts gebildet werden, und zwar in dem Sinne, dass mehr Instanzen der phonologischen Elemente und mehr ausdrucksseitige lexikalische Ketten in einer letztlich verwirrenden Vielgestaltigkeit gebildet werden. Diese Tatsache allein kann aber in einem parallelverarbeitenden System keine Reaktionszeitdifferenzen verursachen. Damit scheidet die Möglichkeit, dass Eigenschaften lexikalischer Ausdrucksseiten an den Frequenzeffekten ursächlich beteiligt sind, aus.

Man muss sich also – in Verstehensrichtung gedacht – um die an die ausdrucksseitige Verarbeitung anschließenden inhaltlichen Prozesse kümmern. Es sind einerseits Mehrdeutigkeiten in Betracht zu ziehen, andererseits aber auch der Einbau eines lexikalischen Inhalts durch zusätzliche Instanzenbildung in mehr oder weniger viele Zusammenhänge, z. B. im Rahmen des Textverstehens. Die Zahl der Mehrdeutigkeiten und der Instanzen insgesamt nimmt naturgemäß zu, wenn ein Wort häufiger im sprachlichen Input erscheint. Wieder gilt aber, dass eine bloße Vermehrung aktivierter oder zu aktivierender Instanzen nicht zu Frequenzeffekten führt, schon gar nicht zu Effekten, bei denen die häufigeren Wörter zeitlich begünstigt erscheinen.

Jetzt gilt aber, dass bei der Verarbeitung von Inhaltsseiten generell inhaltliche Kontexte eine Rolle spielen. Verarbeitungsbahnen werden durch den in Abschnitt 4.4.3 behandelten Kontextfilter ausgewählt. Wenn ein Konzept in mehr Kontexten verankert ist, ist die Chance, dass es rasch aktiviert werden kann und zu weiteren Konsequenzen führt, entsprechend erhöht. Bei nicht passendem Kontext müssen Kontextumschaltungsprozesse stattfinden, wie sie auch zur Erklärung von Priming-Phänomenen angenommen worden sind.

Damit sind aber Frequenzeffekte nicht mehr ausschließlich auf lexikalische Bearbeitungsvorgänge zurückzuführen, sondern auf generell gültige Bedingungen, die vielleicht nicht einmal nur sprachlich sind, sondern mentale Prozesse allgemein betreffen.

4.6.4 Kohorten-Theorie

Die Funktion eines mentalen Lexikons, wie es in den vorangegangenen Kapiteln vorgestellt worden ist, führt tatsächlich zu einem Verhalten in der Sprachperzeption, das in groben Zügen der Kohortentheorie von Marslen-Wilson entspricht, vgl. oben Abschnitt 4.1.2. (Es muss allerdings betont werden, dass das vom Verhalten gilt, nicht von den Details der von Marslen-Wilson angenommen Prozesse und apparativen Voraussetzungen, schon gar nicht, wenn man an das konnektionistische Modell von Gaskell & Marslen-Wilson, 1997, denkt.) Der Verlauf der lexikalischen Vorgänge in der Sprachperzeption führt zur Schrumpfung der Menge der Alternativen, und zwar auch unter Kontexteinfluss, schon ehe das Wort zu Ende ist. Die Weiterverarbeitung des Ergebnisses der lexikalischen Analyse wartet nicht das Erreichen des Wortendes ab.

Dass die Experimente, die der Kohortentheorie zugrunde liegen, einen wichtigen Aspekt der Sprachverarbeitung spiegeln, steht also außer Zweifel. Man sollte sich an dieser Stelle aber zusätzlich überlegen, wie, das heißt durch welche Prozesse, nicht nur aufgrund welcher Gegebenheiten, es zu den in den Verstehensxperimenten festgestellten frühen Reaktionen auf Wörter kommt. Es wäre unplausibel, eine Monitorfunktion anzunehmen, die feststellt, ob zu einem gewissen Zeitpunkt die Kohorte der möglichen Ausdrucksseiten auf einen einzelnen Kandidaten geschrumpft ist. Es gibt keine Vorschläge dafür, wie eine solche Monitorfunktion im Kortex realisiert sein könnte.

Wenn man bedenkt, dass jedes einzelne Phonem eines korrekt lexikalisierten Inputs Konsequenzen in „höheren“ Verarbeitungsbereichen zur Folge hat, kann ein rein ausdrucksseitig definierter „uniqueness point“, also der Punkt, bei dem eine Lautkette eindeutig identifizierbar wird, keine Funktion haben, obwohl selbstverständlich gilt, dass man als Linguist bei einem vorliegenden Lexikon die Lage eines solchen Punkts innerhalb der lexikalischen Ausdrucksseite exakt feststellen kann. Wegen der Redundanz aller neuronalen Repräsentationen ist nicht einmal gewährleistet, dass die Zahl der aktivierten weiterführenden Bahnen ab dem „uniqueness“ point in jedem Fall geringer ist, als zuvor.

Wenn die Aufgabenstellung in einem psycholinguistischen Test so ist, dass ein vereinbartes Target-Wort in gesprochenen Sätzen wiedererkannt werden soll, könnte man sich vorstellen, dass innerhalb des ausdrucksseitigen Lexikons das vereinbarte Wort in irgendeiner Weise ausgezeichnet ist und eine entsprechende Partie des gesprochenen Testsatzes dann diese Ausdrucksseite in besonderer Weise und mit besonderen Konsequenzen aktiviert. Die Kennzeichnung der Ausdrucksseite müsste aber Laut für Laut durch Erwar-

tungspotenziale erfolgen, die ihrerseits durch einen speziellen Mechanismus erzeugt werden müssten, eine Vorstellung, die wohl kaum – und schon gar nicht bei den in den vorangegangenen Kapiteln entwickelten Architekturvorstellungen – akzeptierbar ist. Die zusammenfassende ODER-Zelle, die zur Inhaltsseite zu rechnen ist, ist aufgrund ihrer Charakteristika für eine Kennzeichnung ebenfalls nicht geeignet. Es bleiben als mögliche Ansatzpunkte die Bereiche, die sich in Verstehensrichtung an diese ODER-Zelle anschließen.

In einem symbolverarbeitenden System würde man die Identifikationsleistung u. U. durch einen Vergleich der Lautsequenz des Testsatzes mit dem gespeicherten Target-Wort realisieren. Im Kortex ist ein solcher Vergleich nur schwer denkbar. Es ist ja auch nicht so, dass das Target-Wort während der Lautsequenz des Testsatzes ständig neu aktiviert wird. Was passiert, ist eher so zu beschreiben, dass das Target-Wort erwartet wird und bei seinem Eintreffen dann eine Reaktion erfolgt.

Das Stichwort „Erwartungen“ führt jetzt auf eine mögliche Lösung des Problems. Erwartungen sind sequenzielle Kontexte (vgl. Teil 2, „Grundlagen“, Kap. 2.4.6). Das vereinbarte Target-Wort, das erwartet wird, muss eine episodische Gedächtnisspur bilden, die aus den entsprechenden Kontextbedingungen besteht. Solche episodischen Kontextbedingungen können sich prinzipiell sowohl auf die Ausdrucksseite als auch auf die Inhaltsseite des Target-Worts beziehen. Der Kontext muss regulär aufgefrischt werden, solange die Aufgabe bearbeitet wird. Wenn Input erfolgt, der nicht dem Kontext entspricht, wirkt der Kontext als blockierender Filter. Im positiven Fall dagegen kann aus dem weitergeleiteten Verarbeitungsergebnis eine motorische Reaktion abgeleitet werden, die ja genauso, wie das Target-Wort, zu den vorausgegangenen Vereinbarungen gehört.

4.6.5 Inkrementelle Produktion und Perzeption

Levelt (1989) adaptiert für sein Sprachproduktionsmodell die Idee der inkrementellen Produktion von Kempen & Hoenkamp (1987). Er nimmt auf dem Hintergrund von Fodors Modularitätsprinzip an, dass die Sprachproduktion in einem Zusammenwirken von Spezialisten besteht, die über einen für sie jeweils charakteristischen Input und Output miteinander kommunizieren. Diese Spezialisten können, sofern sie über ein ausreichendes Fragment des Inputs verfügen, zeitparallel arbeiten. Der Grad der Parallelität wird durch zwei Bedingungen bestimmt:

- „A processing component will be triggered by any *fragment* of characteristic input."
- „... this requires that such a fragment can be processed without much lookahead – i. e., that what is done with the fragment should not depend on what will be coming in later fragments."

(Levelt, 1989: 24)

Die zweite Bedingung sagt etwas über die Länge des erforderlichen Fragments und ist ein Hinweis auf eine besondere Eigenschaft dieser Form von Inkrementalität: Das erforderliche Fragment ist zwar minimal zu denken, kann aber unterschiedliche Länge haben. Das scheint auch der Hintergrund der folgenden zusammenfassenden Charakteristik zu sein:

> „*Each processing component will be triggered into activity by a minimal amount of its characteristic input.* In the following chapters we will, time and again, have to consider how small that minimal amount can be." (Levelt, 1989: 26)

Vorausgesetzt wird also, dass ein minimaler vorangegangener, auf einer höheren Ebene angesiedelter Verarbeitungsvorgang abgeschlossen ist, ehe er Konsequenzen auf der nächst niedrigeren Ebene hat. Das muss, wenn die Verarbeitungsvorgänge verschiedene Längen haben, zur Annahme von Zwischenspeichern („buffers") führen, die in dem insgesamt symbolverarbeitenden Kontext kein Fremdkörper sind (vgl. z. B. Levelt, 1989: 414 ff.).

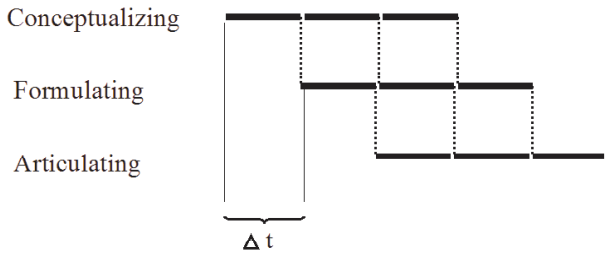

Abbildung 4.6.5–1: Schema zur inkrementellen Produktion, vereinfacht nach Levelt (1989: 25). Alle Verarbeitungsvorgänge haben gleiche Längen.Weitere Erläuterungen im Text.

Die in Abbildung 4.6.5–1 mit Δt bezeichnete Zeitdifferenz entsteht im Wesentlichen durch den Zeitbedarf der Verarbeitungsprozesse auf der höheren Ebene.

Der Vorgang der Lexikonproduktion in der in den Abschnitten 4.4.4 und 4.4.5 entwickelten neuronalen Architektur kann ebenfalls als inkrementell verstanden werden, allerdings in einem anderen Sinn. Eine der Abbildung 4.6.5–1 analoge Darstellung gibt Abbildung 4.6.5–2.

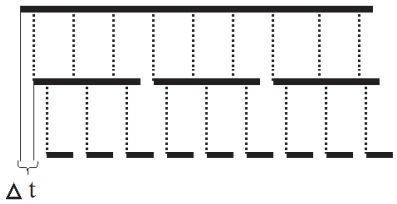

Abbildung 4.6.5–2: Schema zur inkrementellen Produktion in einer neuronalen Architektur.

Der wesentlichste Unterschied zur Position Levelts besteht darin, dass die Weiterverarbeitung auf der nächstniedrigeren Ebene nicht auf den Abschluß eines strukturell definierten Teils der höheren Prozesse wartet. Die Zeitdifferenz Δt wird nur durch die Laufzeit des neuronalen Signals zwischen den Ebenen bestimmt. Der zeitliche Abstand zwischen den Produktionsschritten (wenn man diese Vorstellung beibehalten will) ist auf allen Ebenen gleich und in dem Sinn minimal, dass er den minimalen Schritten auf der untersten Ebene entspricht, die immer die phonologische Ebene ist.

Die Sprachproduktion läuft notwendigerweise zeitlich exakt parallel zu der sie begleitenden Sprachperzeption. Das ist die wesentliche Voraussetzung für ein zeitlich unmittelbares Monitoring des Produktionsprozesses durch eine Watchdog-Instanz (vgl. Teil 2, „Grundlagen“, Abschnitt 2.4.5). Auch dafür sind gleiche Abstände der Verarbeitungsschritte auf allen Verarbeitungsebenen erforderlich. Wenn man die Sprachperzeption einbezieht, entsteht das Schema der Abbildung 4.6.5–3.

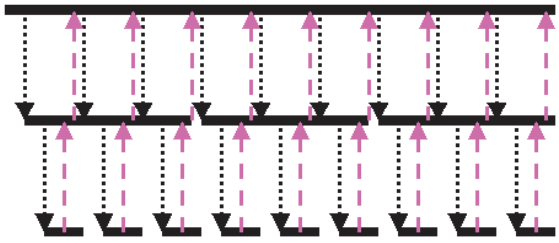

Abbildung 4.6.5–3: Schema zur inkrementellen Produktion mit begleitender Perzeption in einer neuronalen Architektur.

Es ergeben sich Rückwirkungen auf die Arbeit eines „Conceptualizers“, das heißt generell auf den Vorgang des Formulierens. Schließlich müssen auch Versprecher-Phänomene neu interpretiert werden. Diese Gegenstände sind Thema von Teil 7, „Denken und Formulieren“.