

**Manuela Friedrich**

## **Die Entstehung und Überwindung von Übergeneralisierungen im neuronalen Netz ARTMAP**

### **Zusammenfassung**

In dieser Arbeit wird demonstriert, wie begriffliche Übergeneralisierungen in einem neuronalen Netz entstehen und wie sie überwunden werden. Es wurde ein Datensatz genutzt, über dem unterschiedlich abstrakte begriffliche Kategorien gebildet werden können. Jedes Item gehört einem Oberbegriff, einem Basisbegriff und einem Exemplarbegriff an. Während des Lernprozesses im neuronalen Netz ARTMAP wurden einem der beiden Module die begrifflichen Exemplare dargeboten, das zweite Modul erhielt gleichzeitig die zugehörigen Basisbegriffsnamen als Input. Trotz der ausschließlichen Darbietung des jeweils korrekten Namens traten Übergeneralisierungen im Netz auf. Es konnten zwei verschiedene Arten von Übergeneralisierungen identifiziert werden, solche, die beim „Verstehen“ und „Produzieren“ auftreten und solche, die nur beim „Produzieren“ erscheinen. Bei beiden Übergeneralisierungsarten wurde eine Entwicklung von sehr starker Übergeneralisierung über eine Abschwächung bis hin zur regulären Generalisierung beobachtet. Durch das Nachvollziehen dieser Entwicklung unter verschiedenen Bedingungen wurden innerhalb des Modells Mechanismen deutlich, auf deren Basis Übergeneralisierungen verursacht werden. Es zeigte sich, daß insbesondere die Darbietungshäufigkeit in verschiedenen Wirkzusammenhängen agiert und damit zu unterschiedlichen Zeitpunkten unterschiedliche Wirkungen hervorrufen kann. Während die zuerst auftretenden Übergeneralisierungen in Verstehen und Produktion bei sehr selten auftretenden Basisbegriffen am schnellsten überwunden werden, bleiben gerade für diese Exemplare die späteren Übergeneralisierungen beim Produzieren länger bestehen. Weiterhin zeigte sich, daß das „Verhalten“ des Netzes nicht direkt seine interne Verbindungsstruktur widerspiegelt. Diese Differenz erklärt insbesondere das Phänomen „Verstehen vor Produktion“ aus einer Interaktion von Wort- und Begriffsmodul, ohne daß eine zusätzliche Unterscheidung sprachlicher Produktions- und Rezeptionsmodule vorgenommen werden muß.

## Übergeneralisierungen als empirisches Phänomen

Übergeneralisierungen sind ein stabiles empirisches Phänomen, das bei Kindern einer bestimmten Entwicklungsphase häufig beobachtet werden kann und wiederholt in der Literatur beschrieben wurde (z.B. Clark, 1983; Gathercole, 1987; Nelson et al., 1978). Übergeneralisierungen treten in ganz verschiedenen Bereichen auf. Sie liegen dann vor, wenn in der Wahrnehmung oder im Verhalten spezifische Ereignisse die Rolle allgemeiner Ereignisse übernehmen. Die hier untersuchten begrifflichen Übergeneralisierungen sind dadurch charakterisiert, daß Exemplare mit dem Namen eines spezifischen Begriffs, dem sie nicht angehören, benannt werden. Die benannten Exemplare und der spezifische Begriff, dessen Name genutzt wird, gehören dabei einem gemeinsamen übergeordneten Begriff an. Begriffliche Übergeneralisierungen sind also dadurch gekennzeichnet, daß Exemplare eines allgemeinen Begriffes mit dem Namen eines spezifischeren Begriffes bezeichnet werden.

Häufig werden einfache Übergeneralisierungen beobachtet, bei denen *ein* spezifischer Name genutzt wird, um die zu einem allgemeineren Begriff gehörenden Exemplare zu benennen. Diese Art der Übergeneralisierung liegt z.B. vor, wenn der spezifische Name „Hund“ (oder „Wauwau“) auch für andere Tiere genutzt wird. Diese Übergeneralisierung kann unterschiedlich stark ausgeprägt sein. Bei schwächeren Formen werden z.B. nur bestimmte vierbeinige Säugetiere wie Katzen, Schafe oder Ziegen als „Hund“ benannt, bei stärkerer Übergeneralisierung werden z.B. auch Pferde oder Mäuse und sogar Vögel mit dem Namen „Hund“ bezeichnet. Stärkere Übergeneralisierungen werden mit zunehmendem Alter schwächer bis etwa der Zielbereich des Begriffes erreicht ist, womit die Übergeneralisierung überwunden ist.

Im Gegensatz zur einfachen kann man auch mehrfache Übergeneralisierungen beobachten, wenn abwechselnd *verschiedene* spezifische Namen zur Benennung herangezogen werden. Dabei zeigt die Zuordnung der Namen deutlich, daß keine spezifischen Begriffe mit dieser Benennung gemeint sind. So wird ein Fahrrad zu einem Zeitpunkt „Fahrrad“, zu einem anderen „Dreirad“ genannt, ein Dreirad wiederum wird ebenfalls „Fahrrad“ oder „Dreirad“ genannt. Eine Differenzierung von Fahrrad und Dreirad kann weder anhand der Benennung noch anhand des Verhaltens beobachtet werden, beide werden als Exemplare eines gemeinsamen Begriffes behandelt. Diese Art der Übergeneralisierung ist in der Literatur kaum beschrieben worden und wurde von mir bei Lea (ca. 2.2 bis 2.6) beobachtet. Ebenfalls bei Lea (bis ca. 3.3) wurde eine mehrfache Übergeneralisierung beobachtet, die bestimmte Insekten umfaßte. So

benannte sie Bienen, Fliegen, Wespen und Mücken wechselseitig mit einem der vier Namen, wobei die Benennung nur selten und offensichtlich zufällig richtig war.

Die mehrfachen Übergeneralisierungen sind sehr stabil und verschwinden ebenso wie einfache Übergeneralisierungen schrittweise mit dem Etablieren spezifischer Begriffe. Insbesondere verhaltensrelevante Konsequenzen führen zur Differenzierung. So wurden von Lea erst Wespen differenziert, nachdem sie zweimal von einer Wespe gestochen worden war. Wespen und Bienen wurden dann jedoch von ihr als einer Kategorie zugehörig betrachtet und unabhängig von ihrer tatsächlichen Zugehörigkeit mit beiden Namen übergeneralisiert. Analog wurden Mücken und Fliegen wechselseitig korrekt oder falsch benannt. Aus einer komplexen mehrfachen Übergeneralisierung entstanden also zwei zweifache Übergeneralisierungen. Diese verschwanden relativ langsam. Ebenso wurde die zweifache Übergeneralisierung von Fahrrad und Dreirad trotz wiederholter Korrektur relativ lange beibehalten, so lange, bis eine Differenzierung aufgrund des eigenen Dreiradfahrens verhaltensrelevant wurde.

## **Übergeneralisierungen im neuronalen Netz ARTMAP**

Die Simulation des empirischen Phänomens der Übergeneralisierung mit dem sich selbstüberwachendem neuronalen Netz ARTMAP ist deshalb interessant, weil hier die Entstehung, Entwicklung und Überwindung von Übergeneralisierungen direkt nachvollziehbar wird. Es kann sowohl das „Verhalten“ des Netzes beobachtet werden als auch die interne Struktur, auf deren Basis dieses Verhalten entsteht. Innerhalb des Modells kann weiterhin untersucht werden, durch welche Mechanismen Übergeneralisierungen verursacht werden. Dadurch lassen sich Bedingungen identifizieren, die eine Entwicklung von Übergeneralisierungen begünstigen oder verhindern.

Unsere bisherigen Simulationen mit dem nichtüberwachten neuronalen Netz ART1 (Carpenter & Grossberg, 1987), einem Modul des ARTMAP, haben gezeigt, daß das Netz anfänglich sehr allgemeine Kategorien erwirbt. Erst nach mehreren Lernzyklen und in Abhängigkeit von den Parametern Vigilanz, Schwellwert der Merkmalsschicht und Dauer der Resonanz können spezifischere Kategorien gebildet werden. So erklärt sich im Netz eine Übergeneralisierung daraus, daß eine spezifische Benennung zu früh einsetzt, in einer Phase, in der noch keine spezifische begriffliche Kategorisierung möglich ist (Friedrich, 1997).

Um anhand von Simulationen genauer zu untersuchen, unter welchen Bedingungen welche Arten von Übergeneralisierungen auftreten, wie sie überwunden werden und welche Faktoren Einfluß auf diese Entwicklung haben, wurde die vollständige Architektur und Dynamik des ARTMAP (Carpenter und Mitarbeiter, 1991) genutzt. Mit dem ARTMAP ist es möglich, Kategorisierungen über zwei verschiedene Inputtypen miteinander zu assoziieren. Zwei ART1-Module<sup>1</sup> kategorisieren eigenständig ihre Inputs, wobei eines der beiden Module durch das andere überwacht wird. Die eigenständige Kategorisierung des überwachten Moduls erfolgt spontan mit geringer Vigilanz, wird jedoch unter Umständen über eine Vigilanzerhöhung korrigiert. Diese Vigilanzerhöhung wird genau dann ausgelöst, wenn im überwachenden Modul andere als die erwarteten Konsequenzen eintreten, d.h. die bisher erworbene Assoziation zwischen den Modulen nicht bestätigt wird.

Hier wurde das überwachte Modul als Modul der visuell-anschaulichen Kategorisierung (Begriffsmodul) interpretiert, das überwachende dagegen als ein Modul der sprachlichen Kategorisierung (Wortmodul)<sup>2</sup>. Sprachliche Kategorien, d.h. Wörter werden in diesem Fall mit visuellen Kategorien, d.h. sensorischen Begriffen assoziiert, wenn beide Inputs gleichzeitig dargeboten werden und die Module in gemeinsame Resonanz treten. Das sprachliche Modul überwacht das begriffliche Modul, die Benennung beeinflusst damit die begriffliche Differenzierung wesentlich. Werden im Anschluß an einen Lernprozeß mit Benennung weitere Exemplare ohne ihre Namen dargeboten, werden also nur Inputs für das Begriffsmoduls gegeben, dann werden assoziierte Worte von den aktivierten Begriffen geprimt. Dieses Primen kann als eigenständige spontane Benennung der Exemplare durch das Netz interpretiert werden.

Mit dem ARTMAP ist es also möglich, sowohl die Entwicklung der Kategorienstruktur innerhalb der einzelnen ART1-Module (Begriffs- und Wortmodul) als auch die Entwicklung der Verbindungsstruktur zwischen den Modulen (Assoziation von Worten zu Begriffen) zu verfolgen. Zusätzlich kann mit dem ARTMAP das eigenständige Benennungsverhalten des Netzes nach vorangegangener vorgegebener Benennung beobachtet werden.

---

<sup>1</sup> Eine detaillierte Beschreibung der Dynamik des ART1 sowie eine Darstellung der Wirkung von Parametern finden sich in Friedrich (1994) oder in Friedrich und Goede (1997).

<sup>2</sup> Sprachliche Produktion und sprachliche Rezeption sind hier der Einfachheit halber nicht unterschieden.

## **Simulationen**

In den folgenden Simulationen werden Übergeneralisierungen von ihrer Entstehung über ihre Abschwächung bis hin zu ihrer Überwindung nachvollzogen. Dabei werden auch Auswirkungen der Variation von Lernparametern auf das Generalisierungsverhalten des Netzes untersucht.

Es werden jeweils 200 Exemplare gemeinsam mit ihren korrekten Basisnamen zufällig dargeboten. Die Exemplare werden eigenständig vom Begriffsmodul ARTa des ARTMAP kategorisiert, der Basisname des Exemplars wird nichtüberwacht im Wortmodul ARTb verarbeitet. Das ARTb überwacht das ARTa und kann dessen Zuordnung beeinflussen. Das ARTa wiederum kann das ARTb primen. Während des Lernprozesses werden die Herausbildung der Namen (Wortentwicklung), der Aufbau des Begriffssystems sowie die Verbindungsstruktur zwischen Wort- und Begriffsmodul kontinuierlich verfolgt. Weiterhin werden zu bestimmten Zeitpunkten Exemplare ohne ihre Namen dargeboten. Dabei wird beobachtet, welche Kategorien im Wortmodul geprintet werden. Es wird also für verschiedene Entwicklungsstadien geprüft, welche Namen das Netz für die einzelnen Exemplare produziert.

### **Datensatz 1**

Im Datensatz (Tab.1) wurden 8 Exemplare kodiert, die alle vier von 20 Merkmalen gemeinsam haben. Diese stellen die Grundlage für eine gemeinsame Begriffszugehörigkeit aller Exemplare dar und sind deshalb als Oberbegriffsmerkmale gekennzeichnet. In Verbindung mit weiteren acht Merkmalen charakterisieren sie die Basisbegriffszugehörigkeit, während die letzten acht Merkmale spezielle Charakteristika der einzelnen Exemplare darstellen.

Vier Exemplare stimmen zusätzlich in zwei Basisbegriffsmerkmalen überein und werden jeweils gleichzeitig mit dem kodierten Namen „Hund“ dargeboten. Weitere zwei Exemplare gleichen sich in zwei anderen Basisbegriffsmerkmalen und werden mit dem kodierten Namen „Katze“ dargeboten. Die beiden anderen Exemplare haben keine zusätzlichen gemeinsamen Merkmale und werden jeweils mit einem eigenen Namen („Hase“ und „Fuchs“) dargeboten. Jedes Exemplar besitzt weiterhin ein individuelles Merkmal.

Über dem Datensatz können damit vier Basisbegriffe gebildet werden. Das Verhältnis der Auftretenshäufigkeiten ihrer Exemplare ist 4:2:1:1 (HUND:KATZE:HASE:FUCHS).

Für den Datensatz wurde eine Ähnlichkeitsmatrix erstellt, die in Tab.2 zu sehen ist. Als Maß für die Ähnlichkeit wurden das Skalarprodukt zwischen je zwei Exemplar-Vektoren berechnet. Die Ähnlichkeit zu sich selbst beträgt für jedes Exemplar 7, da jeweils 7 Merkmale für ein Exemplar relevant sind. Die Ähnlichkeit zwischen zwei Exemplaren innerhalb der Basisbegriffe beträgt 6. Im Datensatz ist die Ähnlichkeit außerhalb der Basisbegriffe konstant gleich 4. Damit sind alle Basisbegriffe zueinander gleich ähnlich, womit hier der Einfluß unterschiedlicher Ähnlichkeiten auf das Generalisierungsverhalten ausgeschlossen wird.

Tab. 1:           Datensatz:    HUND/ KATZE/ HASE / FUCHS  
 OB-Merkmale: Oberbegriffsmerkmale, BB-Merkmale: Basisbegriffsmerkmale

Items	Exemplar-Merkmale																						
	BB-Merkmale																						
	OB-Merkmale																						
1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
3	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
4	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
5	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
6	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
7	1	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0
8	1	1	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1

Tab. 2:           Ähnlichkeitsmatrix für den Datensatz

Items	1	2	3	4	5	6	7	8
1	7	6	6	6	4	4	4	4
2	6	7	6	6	4	4	4	4
3	6	6	7	6	4	4	4	4
4	6	6	6	7	4	4	4	4
5	4	4	4	4	7	6	4	4
6	4	4	4	4	6	7	4	4
7	4	4	4	4	4	4	7	4
8	4	4	4	4	4	4	4	7

Die Kodierung der Namen erfolgte anhand von 14 phonologischen Merkmalen (Tab.3). Dabei überlagern sich Hase und Katze in zwei Merkmalen, Hase und Hund sowie Fuchs und Hund nur in jeweils einem. Der im Datensatz ebenfalls kodierte Name Tier wurde in den hier beschriebenen Simulationen nicht genutzt. Die Benennung erfolgte also nur auf dem basic level.

Tab. 3: Namensdatensatz: Hund/ Katze/ Hase/ Fuchs/ (Tier)

<i>Silbenanfangsrand</i>				<i>Kern</i>			<i>Silbenendrand bzw. 2. Silbe</i>							
<i>h</i>	<i>k</i>	<i>f</i>	<i>t</i>	<i>ʊ</i>	<i>a<sup>3</sup></i>	<i>i</i>	<i>ʀ</i>	<i>n</i>	<i>z</i>	<i>t<sup>s</sup></i>	<i>ks</i>	<i>t</i>	<i>ə</i>	<i>Name</i>
1	0	0	0	1	0	0	0	1	0	0	0	1	0	Hund
0	1	0	0	0	1	0	0	0	0	1	0	0	1	Katze
1	0	0	0	0	1	0	0	0	1	0	0	0	1	Hase
0	0	1	0	1	0	0	0	0	0	0	1	0	0	Fuchs
0	0	0	1	0	0	1	1	0	0	0	0	0	0	Tier

Für die Darbietung der Items wurde eine Zufallsreihenfolge erzeugt und aus Gründen der Vergleichbarkeit für alle Parameterkombinationen konstant gehalten. Zu Kontrollzwecken wurden die Simulationen außerdem mit unterschiedlichen Zufallsreihenfolgen wiederholt.

### Wortentwicklung

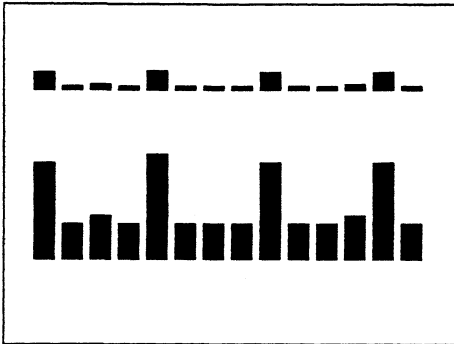
Zum Erlernen der Namenskategorien wurden für das ARTb eine konstante Resonanzdauer (k<sub>kb</sub>=3) sowie eine konstante Vigilanz (p<sub>b</sub>=0.7) gewählt.

Nach der Darbietung von 10 Items hat das Netz zwei Namenskategorien erworben, eine weitere wird unmittelbar mit dem 10. Item konstituiert. Die zuerst gelernte Kategorie kodiert den Namen „Hund“ (Abb.1a), die zweite stellt eine unscharfe Überlagerung der Namen „Katze“ und „Hase“ dar, bei der die überlappenden Merkmale im Vordergrund stehen (Abb.1b). Die ab dem 10. Item konstituierte Namenskategorie „Fuchs“ wird während der nächsten 10 Items gefestigt (Abb.1c). Die Überlagerung der Namen „Katze“ und „Hase“ innerhalb einer Kategorie besteht zu diesem Zeitpunkt weiter (Abb.1d). Schließlich differenziert das Netz während der 10 nächsten Items den Namen „Hase“ (Abb.1e), so daß sich die gemischte Kategorie zum Namen „Katze“ entwickelt (Abb.1f). Nach 30 Items sind alle vier Namen vom Netz erworben.

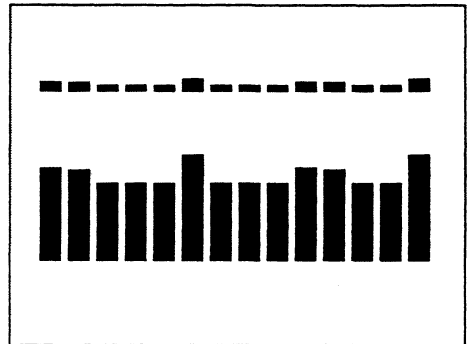
<sup>3</sup> a ist der untere Zentralvokal

Abb.1a-f: Entwicklung der Namenskategorien über dem Namensdatensatz:  
Gewichtestruktur der phonetischen Merkmale für die Namenskategorien,  
bottom up-Gewichte (Prototypen): oben in den jeweiligen Abbildungen,  
top down-Gewichte (Erwartungen): unten in den jeweiligen Abbildungen

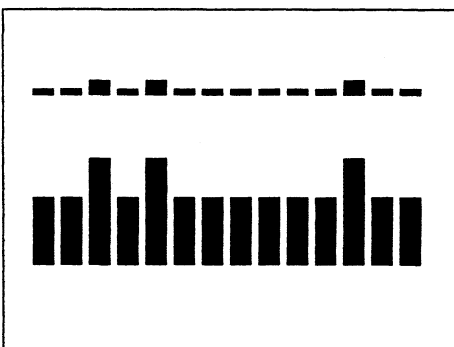
a) Kategorie „Hund“ nach 10 Items



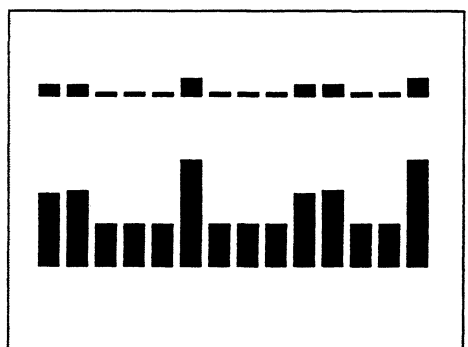
b) Mischkategorie nach 10 Items



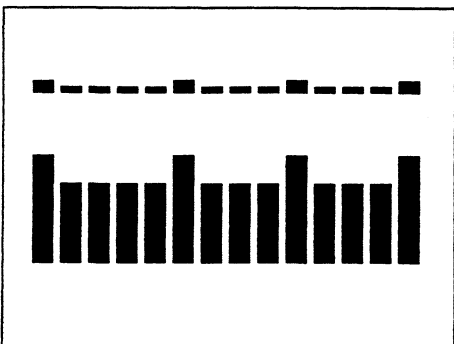
c) Kategorie „Fuchs“ nach 20 Items



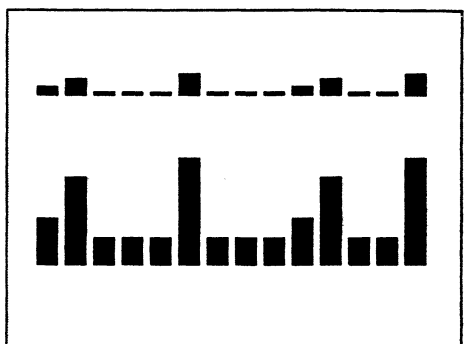
d) Mischkategorie nach 20 Items



e) Kategorie „Hase“ nach 30 Items



f) Kategorie „Katze“ nach 30 Items





## Entwicklung des Begriffssystems

Die Entwicklung des Begriffssystems wurde kontinuierlich während der Zuordnung benannter Exemplare verfolgt. Es wurden Simulationen mit unterschiedlicher Resonanzdauer für das ARTa ( $kka=1$  und  $kka=3$ ) durchgeführt. Weiterhin wurde die Stärke des Lernens der intermodularen Verbindungen in drei Stufen variiert ( $kkm=1/3/5$ ). Es werden exemplarisch drei Simulationen detailliert beschrieben.

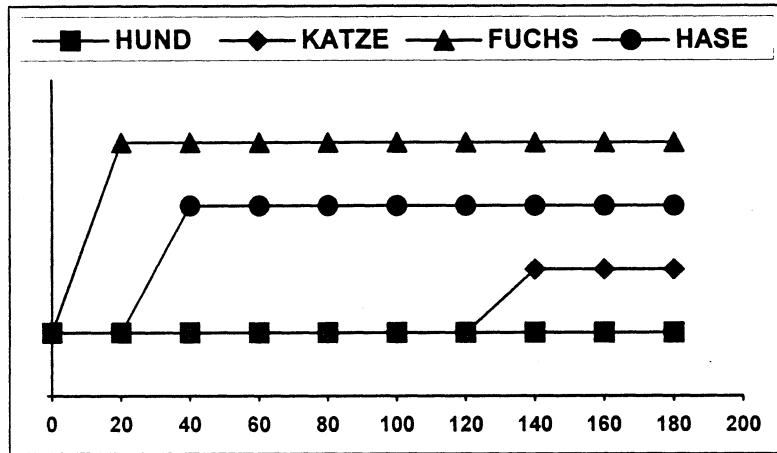
Unter den verwendeten Parameterkonstellationen bildete sich in allen Simulationen anfänglich eine allgemeine Kategorie heraus, der alle Exemplare zugeordnet werden. Dies entspricht einer starken Übergeneralisierung. Nach der Darbietung von 140 Exemplaren entstanden in jedem der im folgenden beschriebenen Fälle vier begriffliche Kategorien, die jeweils mit ihrem korrekten Basisnamen verbunden waren. Damit werden ab diesem Zeitpunkt bei entsprechender Benennung alle Exemplare regulär generalisiert.

Obwohl sich Anfangs- und Endzustand des Begriffssystems unter allen drei Bedingungen gleichen, verlief die Differenzierung der einzelnen Basisbegriffe qualitativ und quantitativ unterschiedlich. Die Unterschiede äußerten sich insbesondere in der zeitlichen Abspaltung einzelner Exemplare von der allgemeinen Kategorie und damit in der Reihenfolge des Konstituierens spezifischer Begriffe. Weiterhin bildeten sich in unterschiedlicher Weise Kategorien eines mittleren Abstraktionsgrades, d.h. solche, denen die Exemplare von zwei oder drei, nicht jedoch von allen vier Basisbegriffen zugeordnet wurden.

Im Falle gleicher Resonanzdauer von Wort- und Begriffsmodul sowie der Verbindung zwischen beiden ( $kka=kkb=kkm=3$ , dargestellt in Abb.2a) bildete sich bereits nach wenigen Exemplaren eine Kategorie heraus, der das Fuchs-Exemplar zugeordnet wird. So existiert nach 10 Items eine Kategorie FUCHS und nach weiteren 20 Items auch eine Kategorie HASE. Es sind also zuerst zwei spezifische Kategorien für die seltener vorkommenden Exemplare entstanden. Die häufiger vorkommenden Hunde- und Katzen-Exemplare werden noch über einen relativ langen Zeitraum der zuerst gebildeten allgemeineren Kategorie zugeordnet. Erst nach 140 Items werden die gegenüber den Hunde-Exemplaren selteneren Katzen-Exemplare abgespalten und konstituieren eine eigene Kategorie KATZE, so daß der anfänglich allgemeinen Kategorie nun nur noch die Hunde-Exemplare zugeordnet werden. In dieser Simulation spalten sich also nacheinander Exemplare einzelner Basisbegriffe von dem allgemeinen Begriff ab und bilden jeweils eigene Kategorien. Die allgemeine Kategorie entwickelt sich schließlich zu der spezifischen Kategorie des am häufigsten benannten und übergeneralisierten Basisbegriffs.

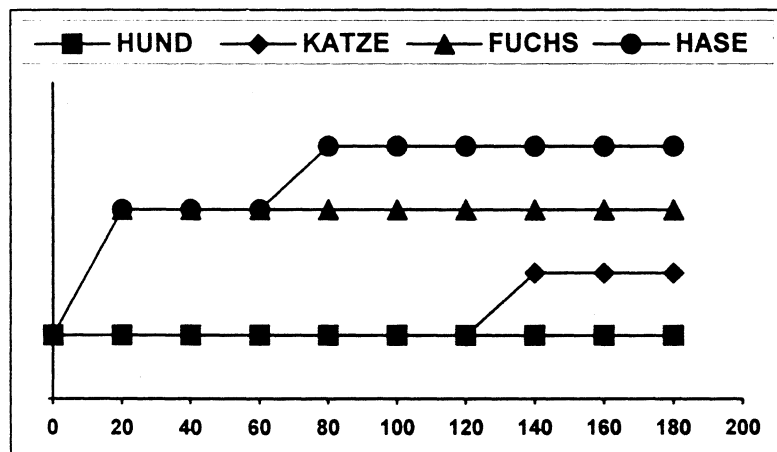
Abb.2a-c: Differenzierung des Begriffssystems mit zunehmender Itemdarbietung unter verschiedenen Parameterbedingungen (a, b und c)

a)  $kka=3, kkb=3, kkm=3$



Bei einem verringertem Lernen innerhalb des Begriffsmoduls ( $kka=1$ ) entstehen aus einer anfänglich starken Übergeneralisierung zwei schwächere Übergeneralisierungen (Abb.2b). So wird auch hier vom Fuchs-Exemplar relativ früh eine spezifische Kategorie konstituiert, dieser wird dann auch das Hasen-Exemplar zugeordnet. Nach 80 Items werden Fuchs- und Hasen-Exemplare aufgrund der Bildung einer Hasen-Kategorie differenziert, und nach 140 Items werden Katzen und Hunde über das Konstituieren einer Katzen-Kategorie begrifflich getrennt.

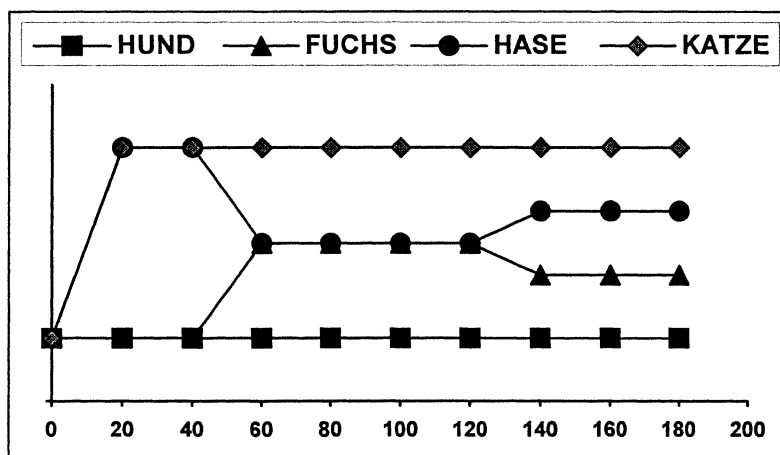
b)  $kka=1, kkb=3, kkm=3$



Verringertes Lernen innerhalb des Begriffsmoduls verbunden mit stärkerem Lernen der Verbindungen zwischen den Modulen führt zu einer etwas anderen begrifflichen Differenzierung (Abb.2c). Hier wird vom Hasen eine erste spezifische Kategorie konstituiert, der dann auch

die Katzen-Exemplare zugeordnet werden. Nachdem vom Fuchs ebenfalls eine spezifische Kategorie gebildet wurde (etwa nach 50 Items), verläßt der Hase die Kategorie HASE/KATZE und wird der Fuchs-Kategorie zugeordnet. Damit gibt es neben der gemeinsamen Kategorie FUCHS/HASE, die sich erst nach 140 Items aufspaltet, bereits eine Kategorie KATZE und eine Kategorie HUND. Für Hunde- und Katzen-Exemplare existieren also schon relativ früh zwei getrennte Kategorien. Die zeitige Differenzierung von Hunden und Katzen ist jedoch nicht auf eine eigenständige Bildung ihrer spezifischen Kategorien zurückzuführen, sondern ist Ergebnis des Ablösens der Hasen- und Fuchs-Exemplare von den jeweiligen allgemeineren Kategorien. Es zeigt sich auch hier, daß bei korrekter Benennung für den Fall gleicher zwischenbegrifflicher Ähnlichkeit die Exemplarhäufigkeit der einzelnen Begriffe bedeutsam für die begriffliche Differenzierung ist. Seltener Exemplare werden eher von der allgemeinen Kategorie abgelöst und bilden dadurch früher eine eigene spezifische Kategorie.

c)  $kka=1, kkb=3, kkm=5$

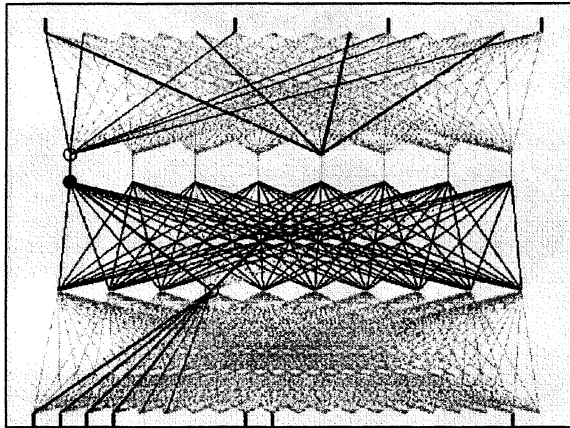


### Entwicklung der Verbindungsstruktur

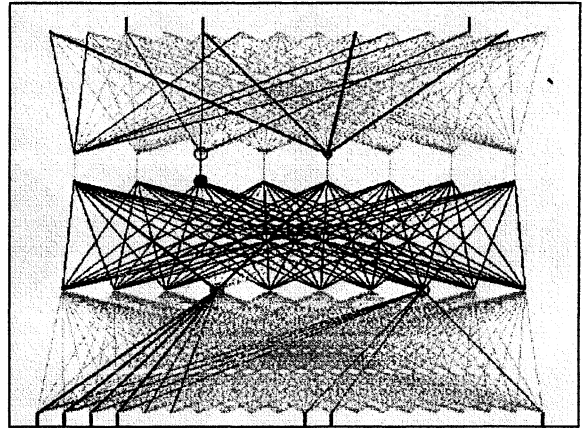
Im Falle gleicher Resonanzdauer von Wort- und Begriffsmodul sowie der Verbindung zwischen beiden ( $kka=kkb=kkm=3$ ) sind nach Darbietung und korrekter Benennung von 5 Items die beiden zuerst gebildeten Wortkategorien „Hund“ und „Katze/Hase“ mit dem einen allgemeinen Begriff verbunden, dem anfänglich alle Exemplare zugeordnet werden (Abb.3a). Unmittelbar mit dem Konstituieren der Namenscategory „Fuchs“ wird die Bildung einer begrifflichen Kategorie FUCHS initiiert, die mit dem Namen „Fuchs“ verbunden ist (Abb.3b). Nach 30 Items existiert zusätzlich eine Verbindung zwischen den neu erworbenen Namenscategory „Hase“ und der dadurch gebildeten begrifflichen Kategorie HASE (Abb.3c).

Abb.3a-f: Entwicklung der Verbindungsstruktur innerhalb und zwischen den Modulen  
 oben: Gewichte im ARTb-Wortmodul (mit  $k_{kb}=3$  erworben),  
 unten: Gewichte im ARTa-Begriffsmodul (mit  $k_{ka}=3$  erworben),  
 Mitte: Gewichte zwischen den Modulen (mit  $k_{km}=3$  erworben).

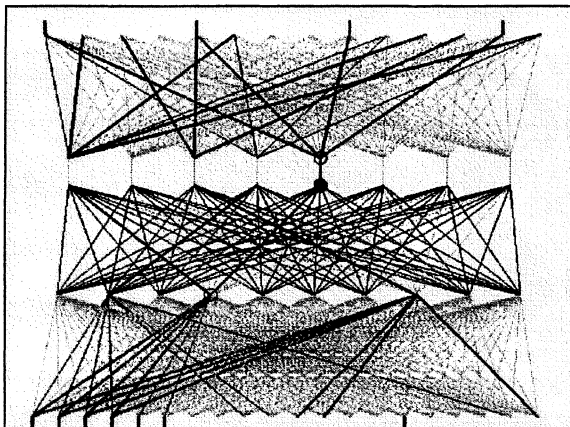
a) Verbindungen nach 5 Items



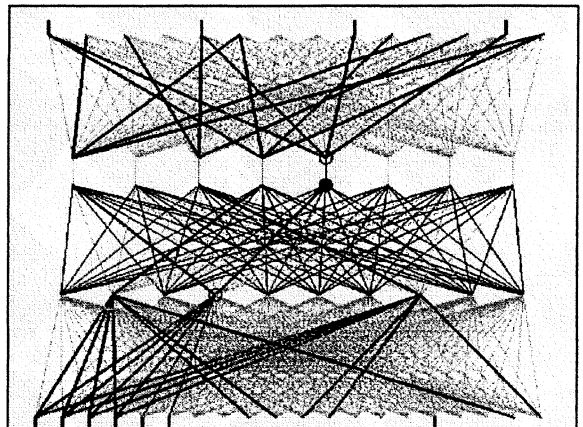
b) Verbindungen nach 10 Items



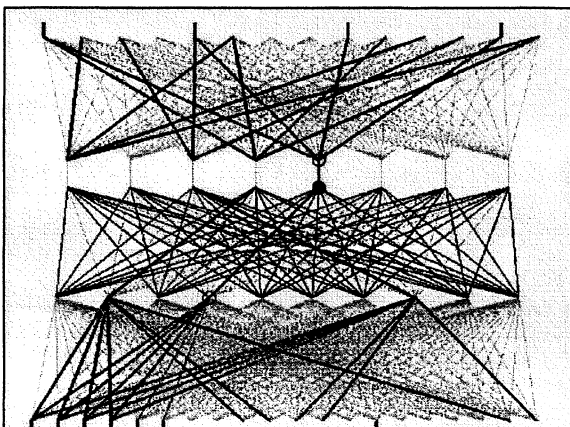
c) Verbindungen nach 30 Items



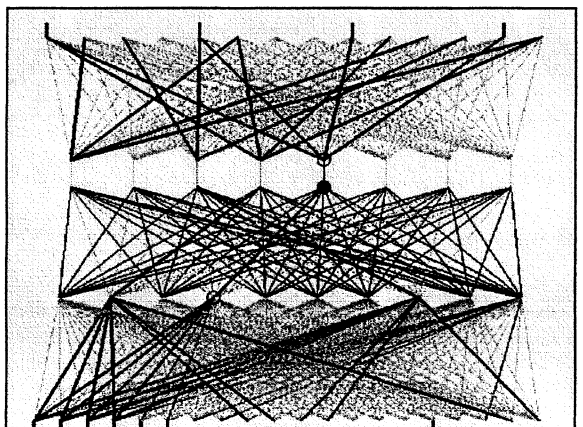
d) Verbindungen nach 60 Items



e) Verbindungen nach 80 Items



f) Verbindungen nach 140 Items



Dieser Zustand des Vorhandenseins einer 2:1-Verbindung und zweier 1:1-Verbindungen ändert sich über einen langen Zeitraum nur quantitativ, d.h. in den Verbindungsstärken (Abb.3c/d/e). Dabei wird die Verbindung vom Namen „Katze“ zur gemeinsamen begrifflichen Kategorie HUND/KATZE aufgrund der selteneren Assoziation so geschwächt, bis sie völlig gelöst ist. Das ermöglicht die Bildung einer neuen begrifflichen Kategorie KATZE, die mit dem Namen „Katze“ verbunden wird, so daß nach ca. 140 Items vier stabile 1:1-Verbindungen im Netz vorhanden sind (Abb.3f).

Unter der Bedingung, daß im Begriffsmodul langsamer gelernt wird, entwickelt sich die Verbindungsstruktur anfangs ebenso. Nach 30 Items wird jedoch der Name „Hase“ zusätzlich zum Namen „Fuchs“ mit der begrifflichen Kategorie FUCHS/HASE verbunden. Damit existieren jetzt im Netz zwei 2:1-Verbindungen und keine 1:1-Verbindung. Nach 70 Items wird die Verbindung des Namens „Hase“ zum gemeinsamen Begriff FUCHS/HASE gelöst, was zum Konstituieren der begrifflichen Kategorie HASE und zur Herstellung von zwei 1:1-Verbindungen führt. Analog wird nach etwa 140 Items die 2:1-Verbindung von „Hund“ und „Katze“ zu ihrem gemeinsamen Begriff aufgehoben, wodurch ebenfalls zwei 1:1-Verbindungen der Basisbegriffe zu ihren jeweiligen Namen erworben werden. Damit sind auch hier nach 140 Items alle vier Basisbegriffe mit ihren entsprechenden Namen verbunden.

Durch stärkeres Lernen der Verbindungen zwischen den Modulen entstehen nach 20 Items zwei 1:1-Verbindungen der Begriffe HUND/HASE und KATZE/FUCHS zu den Namen „Hund“ und „Katze/Hase“. Letzterer entwickelt sich innerhalb der folgenden 10 Items zum Namen „Katze“. Die Namen „Fuchs“ und „Hase“ sind in diesem Stadium nicht mit dem Begriffsmodul verbunden und werden demzufolge vom Netz nicht „verstanden“. Nach etwa 50 Items wird eine 2:1-Verbindung des gemeinsamen Begriffs FUCHS/HASE zu beiden Namen erworben. Aufgrund der gleichen Häufigkeit der Fuchs- und Hasenexemplare bleibt diese Verbindungsart über einen langen Zeitraum erhalten. Erst nach der Darbietung von ca. 130 Items wurde die Verbindung zum Namen „Hase“ aufgrund lokaler Häufigkeitsschwankungen so geschwächt, daß eine neue Kategorie HASE entstehen kann, die mit dem Namen „Hase“ verbunden wird. Damit wurden in diesem Fall auch nach 140 Items vier 1:1-Verbindungen für die Basisbegriffe und ihre Namen gebildet. Insbesondere bei stärkerem Lernen zwischen den Modulen ist das System jedoch sehr sensibel gegenüber Änderungen in der Darbietungsreihenfolge. So kann die 2:1-Verbindung von FUCHS/HASE zu ihren Namen bei einer anderen Zufallsreihenfolge der Items auch bedeutend länger bestehen bleiben.

## Primen der Namen

Im ersten untersuchten Fall, bei dem die Resonanzdauer innerhalb und zwischen den Modulen identisch ist ( $kka=kkb=kkm=3$ , dargestellt in Tab.4, Spalte 1), wird nach der Darbietung von 20 korrekt benannten Exemplare aufgrund der Zuordnung zur allgemeinen Kategorie von allen Exemplaren die Wortkategorie geprint, die eine Überlagerung der Namen Katze und Hase  $[(k/h)a(ts/z)\partial]$  darstellt. Damit ordnet das Netz den Exemplaren noch keinen korrekten Namen zu, es befindet sich in einer „Babbelphase“. Nach weiteren 20 benannten Exemplaren werden ohne Benennung wieder alle Exemplare der allgemeinen Kategorie zugeordnet. Diese ist aktuell am stärksten mit der Namenskategorie „Hund“ verbunden. Damit primen alle acht Exemplare den Namen „Hund“, obwohl in dieser Entwicklungsphase bereits zwei spezifische Kategorien (FUCHS und HASE) vom Netz erworben wurden, die jeweils mit ihren richtigen Namen verbunden sind (vgl. Abb.2a und Abb.3c). Obwohl also das Netz spezifische Namen für die Basisbegriffe schon „verstehen“, d.h. die Exemplare als Basisbegriffe erkennen kann, produziert es eigenständig nur den Namen eines einzigen Basisbegriffes für alle Exemplare. Diese starke Übergeneralisierung ist also nicht qualitativ in der Struktur der Verbindungen des Netzes begründet, sondern quantitativ in der Stärke der bereits bestehenden Verbindungen innerhalb des Begriffsmoduls.

Nach insgesamt 60 Items ist die Kategorie FUCHS bereits so gut gelernt, daß ihr die Fuchs-Exemplare spontan zugeordnet werden. Damit wird der Name „Fuchs“ korrekt geprint. Die anderen Exemplare werden weiterhin der zuerst gebildeten Kategorie zugeordnet. Aufgrund von lokalen Häufigkeitsschwankungen in der Darbietung des Datensatzes ist zu diesem Zeitpunkt der Name „Katze“ stärker mit der allgemeinen Kategorie verbunden als der Name „Hund.“ Das führt dazu, daß Hunde-, Katzen- und Hasen-Exemplare den Namen „Katze“ primen. Nach weiteren 20 Items ist die Verbindung zum Namen „Hund“ wieder stärker, so daß außer dem Fuchs alle Exemplare die Benennung „Hund“ auslösen. Damit tritt deutlich eine mehrfache (hier zweifache) Übergeneralisierung in Erscheinung. Zu verschiedenen Zeitpunkten werden unterschiedliche spezifische Namen verwendet, ohne daß die Exemplare wirklich spezifischen Kategorien zugeordnet werden. Einzig durch Schwankungen in der Stärke der Verbindungen einer allgemeinen Kategorie zu assoziierten spezifischen Namen wird über das Benennungsverhalten der Anschein erweckt, daß bereits mehrere spezifische Kategorien angesteuert werden.

Tab. 4: Namenszuordnung beim Primen für unterschiedliche Parameter zu verschiedenen Zeiten

Anzahl der Items	kka=3, kkb=3, kkm=3	kka=1, kkb=3, kkm=3	kka=1, kkb=3, kkm=5	Generalisierung	
20	Katze-Hase (alle)	Katze (alle)	Hund (alle)	starke Übergenera- lisierung  ↓	
40	Hund (alle)	Hund (alle)			
60	Katze (HU,KA,HA) Fuchs (FU)	Katze (alle)	Hund (HU, HA,FU) Katze (KA)		
80	Hund (HU,KA,HA) Fuchs (FU)				
100	Hund (HU, KA) Fuchs (FU) Hase (HA)	Hund (alle)			
120			Hund (HU, FU) Katze (KA) Hase (HA)		
140					
160					
180					
200	Hund (HU) Fuchs (FU) Hase (HA) Katze (KA)	Hund (alle)			
220			Hund (HU) Katze (KA) Hase (HA) Fuchs (FU)	schwächere Übergenera- lisierung  ↓	
240					
260		Hund (HU,KA,HA) Fuchs (FU)			
280		Hund (HU, KA) Fuchs (FU) Hase (HA)			
300		Hund (HU) Fuchs (FU) Hase (HA) Katze (KA)			
					reguläre Generalisierung

Nach der benannten Darbietung von insgesamt 100 Items wird auch der Hase seinem Basisbegriff zugeordnet und vom Netz korrekt benannt. Damit besteht die Übergeneralisierung nur noch für die Katzen-Exemplare. Die Bewichtung der Verbindungen von der allgemeinen Kategorie zu den spezifischen Namen hat sich jetzt aufgrund der globalen Häufigkeiten im Datensatz stabilisiert. Aus diesem Grund wird von Hunden- und Katzen-Exemplaren im folgenden nur noch der Name „Hund“ geprint. Obwohl das Netz bereits nach 140 Items bei benannter Darbietung von Katzen-Exemplaren diese dem Basisbegriff KATZE zuordnet, bleibt die Übergeneralisierung beim Primen noch über einen Zeitraum von 60 Items bestehen. Erst nach Darbietung von insgesamt 200 Exemplaren mit korrekter Benennung werden auch eigenständig vom Netz allen Exemplaren die richtigen Namen ihrer Basisbegriffe zugewiesen. Es treten jetzt keine Übergeneralisierungen mehr auf.

Im Falle eines verringerten Lernens im Begriffsmodul verläuft die Entwicklung prinzipiell gleich, die einzelnen Phasen verschieden starker Übergeneralisierungen bleiben jedoch länger erhalten (Tab.4, Spalte 2). So tritt eine ausgeprägte Phase mit zweifacher Übergeneralisierung der Namen „Katze“ und „Hund“ auf. Weiterhin wird nach 200 Items immer noch der Name „Hund“ von allen Exemplaren geprint, obwohl bereits alle vier Basisbegriffe existieren und mit ihren Namen verbunden sind. Nach 260 Items wird der Fuchs korrekt benannt, nach 280 Items auch der Hase. Nach 300 Items schließlich print das Netz alle Namen korrekt. Es zeigt sich hier deutlich, daß die Zeitspanne von der Bildung des jeweiligen Basisbegriffs bis zur korrekten Benennung seiner Exemplare variiert. Diese Differenz ist sowohl für den Fuchs (240) als auch für den Hasen (200) größer als für die Katzen (160). Obwohl der Begriff KATZE sehr spät erworben wurde, benötigt er weniger Zeit, um gefestigt zu werden, was an der häufigeren Darbietung der Katzen-Exemplare liegt. Daran wird deutlich, daß sich nach dem Konstituieren der spezifischen Kategorie der Einfluß der Häufigkeit auf das Generalisierungsverhalten gerade umkehrt.

Bei stärkerem Lernen zwischen den Modulen (Tab.4, Spalte 3) werden nach 60 Items zuerst die Katzen-Exemplare richtig benannt, während alle anderen Exemplare den Namen „Hund“ primen. Nach weiteren 60 Items print das Netz auch den Namen „Hase“ korrekt über die Kategorie HASE/FUCHS, obwohl mit dieser ebenfalls der Name „Fuchs“ verbunden ist. Vom Fuchs wird dagegen spontan die besser gelernte Kategorie HUND aktiviert, so daß weder der Name „Fuchs“ noch der Name „Hase“ sondern der Name „Hund“ geprint wird. Insgesamt werden 220 korrekte Benennungen benötigt, um den Entwicklungszustand einer regulären Generalisierung zu erreichen.



## Ergebnisse der Simulationen

Insgesamt zeigen die Simulationen, daß trotz ausschließlich korrekter Benennung der Exemplare einfache und mehrfache Übergeneralisierungen im Netz auftreten. Diese sind darin begründet, daß die spezifische Benennung zu früh einsetzt, in einer Phase, in der das Begriffsmodul noch keine spezifischen Kategorien erwerben oder die bereits erworbenen spezifischen Kategorien noch nicht eigenständig aktivieren kann.

Die Unfähigkeit zur frühen begrifflichen Differenzierung liegt in der Verbindungsstruktur zwischen Wort- und Begriffsmodul begründet: Bei langsamem Lernen können mehrere Namen zu einer allgemeinen Kategorie assoziiert werden. Erst wenn aufgrund häufigerer alternativer Benennung die Verbindung eines spezifischen Namens zu einer allgemeinen Kategorie gelöst wurde, kann ebendiese spezifische Benennung eine Vigilanzerrhöhung auslösen und damit die Bildung einer spezifischen begrifflichen Kategorie induzieren. Eine weitere Ursache für die fehlende begriffliche Differenzierung kann der Lernzustand der allgemeinen Kategorie sein: Wenn die allgemeine Kategorie noch schlecht gelernt und damit sehr unscharf ist, kann selbst eine maximale Vigilanzerrhöhung keine Ablösung von ihr bewirken (Friedrich, 1997). Der Grund für die Unfähigkeit zur spontanen Aktivierung einer bereits konstituierten spezifischen Kategorie ist ebenfalls im Lernzustand des Begriffsmoduls zu finden: Eine allgemeine Kategorie wird zuerst gebildet und ist daher besser gelernt als später erworbene spezifische Kategorien. Demzufolge erhält die allgemeine Kategorie stärkere Signale, gewinnt die Konkurrenz und wird spontan aktiviert. Nur durch eine Vigilanzerrhöhung kann dieser Vorrang überwunden werden.

Entsprechend dieser unterschiedlichen Ursachen kann man in den Simulationen auch zwei verschiedene Übergeneralisierungsformen identifizieren. Die beiden Formen folgen für ein einzelnes Ober-/Basisbegriffspaar zeitlich aufeinander, betrachtet man jedoch Exemplare unterschiedlicher Basisbegriffe, so können sie auch gleichzeitig auftreten.

Bei der zeitlich ersten Übergeneralisierungsform befindet sich das Netz in einem Zustand, in dem ein allgemeiner Begriff herausgebildet wurde, während spezifische Kategorien für die übergeneralisierten Exemplare noch nicht existieren. Die übergeneralisierten Exemplare werden alle dem allgemeinen Begriff zugeordnet, der mit mehreren spezifischen Namen verbunden ist. Einer dieser spezifischen Namen wird von der allgemeinen Kategorie geprint, so daß er für Exemplare anderer spezifischer Begriffe übergeneralisiert wird. Der übergeneralisierte Name kann unter bestimmten Bedingungen wechseln, es werden dann zu unterschiedlichen

Zeitpunkten verschiedene spezifische Namen geprint. Bei dieser Übergeneralisierungsform ist ein „Verstehen“ spezifischer Namen als spezifische Begriffszugehörigkeit eines Exemplars nicht möglich, da die entsprechenden Exemplare in jedem Fall nur als einer allgemeinen Kategorie zugehörig erkannt werden. Ebenso wenig möglich ist die korrekte „Produktion“ spezifischer Namen für spezifische Begriffe.

Die andere Übergeneralisierungsform ist nicht in der Verbindungsstruktur zwischen Wort- und Begriffsmodul begründet, sondern liegt ausschließlich am Entwicklungszustand der Kategorien im Begriffssystem. Sie tritt nur bei eigenständiger Produktion von Namen in Erscheinung, wenn bereits spezifische begriffliche Kategorien konstituiert wurden, diese jedoch noch nicht hinreichend gelernt sind. Die spezifischen Kategorien sind hier jeweils mit ihren eigenen Namen verbunden. Bei korrekter Benennung erfolgt eine korrekte Zuordnung zu den jeweiligen Kategorien. Damit ist ein „Erkennen“ bzw. „Verstehen“ möglich. Ohne Benennung werden jedoch bei niedriger Vigilanz viele Exemplare der zuerst konstituierten allgemeineren Kategorie zugeordnet, da diese bereits besser gelernt ist und spontan die Konkurrenz gewinnt. Dabei wird der mit dieser Kategorie verbundene spezifische Name übergeneralisiert. In dieser Entwicklungsphase kann das Netz die spezifische Begriffszugehörigkeit also schon „verstehen“, wenn ein spezifischer Name darauf hinweist, jedoch kann der spezifische Name eigenständig noch nicht als korrekte Benennung von Exemplaren „produziert“ werden.

Bei beiden Übergeneralisierungsformen treten verschiedene Grade in Erscheinung. Die Entwicklung verläuft jeweils von starker Übergeneralisierung über eine Abschwächung bis hin zur regulären Generalisierung. Als relevant für die Übergeneralisierung verschiedener Exemplare hat sich ihre Häufigkeit herausgestellt. Selten auftretende Exemplare bilden früher eine eigene Kategorie und werden beim Verstehen nicht so lange übergeneralisiert. Dafür gibt es zwei Gründe, die bei unterschiedlichen Parameterkonstellationen in Erscheinung treten. Zum einen verleihen die häufigeren Exemplare der allgemeinen Kategorie eine bestimmte Nuancierung, so daß die selteneren dadurch untypischer werden. Vor allem wird jedoch die Verbindung der allgemeinen Kategorie zu den Namen der seltenen Exemplare früh gelöst, was bei der Benennung der entsprechenden Exemplare eine Vigilanzhöhung im Begriffsmodul auslöst. Bei stärkerem Lernen zeigt sich aber der Effekt, daß anstelle der globalen Exemplarhäufigkeit die lokale Häufigkeit in einem bestimmten Zeitabschnitt entscheidend wird. Damit können hier auch häufigere Exemplare relativ früh eigene Kategorien bilden. Weiterhin tritt nach dem Konstituieren einer spezifischen Kategorie ein gegenteiliger Effekt der Häufigkeit

auf. Die spezifischen Kategorien werden durch häufigere Exemplare schneller gefestigt, wodurch die Übergeneralisierung beim Produzieren schneller überwunden wird.

Die hier variierten Lernparameter führen zu ähnlichen Entwicklungsverläufen im Generalisierungsverhalten. Unterschiede finden sich in der Geschwindigkeit dieses Prozesses, in der Robustheit gegenüber Änderungen der Darbietungsreihenfolge bzw. gegenüber lokalen Häufigkeitsschwankungen des Inputs sowie in der konkreten Generalisierung bestimmter Exemplare. So können einerseits spezifische Kategorien sukzessive von der allgemeinen abgelöst werden, wodurch die Übergeneralisierung schwächer wird und die allgemeine Kategorie sich zu einer spezifischen entwickelt. Andererseits können sich verschiedene Kategorien eines mittleren Abstraktionsgrades bilden, was zur Aufspaltung einer starken Übergeneralisierung in mehrere schwächere führen kann. Schließlich entwickeln sich jedoch alle Kategorien zu dem mit der Benennung angezielten Basisniveau.

Als wesentliches Ergebnis der Simulationen muß hervorgehoben werden, daß das eigenständige Benennungsverhalten des Netzes nicht unmittelbar seine innere Verbindungsstruktur reflektiert. Eine einzige allgemeine Kategorie kann zu unterschiedlichen Zeitpunkten unterschiedlich benannt werden und damit den Anschein einer begrifflichen Differenzierung erwecken. Umgekehrt kann das Begriffssystem trotz bestehender Übergeneralisierung in der Benennung bereits stark differenziert sein. Dies impliziert für empirische Untersuchungen die Notwendigkeit einer besonders sorgfältigen Herangehensweise bei der Erkundung des kindlichen Umgangs mit unterschiedlich abstrakten Begriffen.

## Literatur

- Carpenter, G.A. & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37, 54-115.
- Carpenter, G.A., Grossberg, S. & Reynolds, J. H. (1991). ARTMAP: supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, 4, 565-588.
- Clark, E.V. (1983). Convention and contrast in acquiring the lexicon. In T. B. Seiler & W. Wannemacher (Hrsg), *Concept development and the development of word meaning*. New York: Springer Verlag.
- Friedrich, M. (1997). Der Erwerb hierarchisch ordenbarer Kategorien in einem neuronalen Netz der Adaptiven Resonanztheorie. in diesem Band.
- Friedrich, M. (1994). Modellierung und Simulation kategorialer Strukturbildung - Eine Anwendung der Adaptiven Resonanztheorie auf die Begriffsbildung. Dissertation.
- Friedrich, M. & K. Goede (in Vorbereitung). Übergeneralisierungen: Empirische Phänomene und Simulationen im neuronalen Netz ARTMAP. Zur Veröffentlichung in der *Zeitschrift für Psychologie* vorgesehen.
- Friedrich, M. & Goede, K. (1997). Eine Anwendung der Adaptiven Resonanztheorie auf den Erwerb und die Repräsentation begrifflichen Wissens. *Zeitschrift für Psychologie*, 205, 1-35.
- Gathercole, V. (1987). The contrastive hypothesis for the acquisition of word meaning: A reconsideration of the theory. *Journal of Child Language*, 14, 493-531.
- Goede, K. (1989). Möglichkeiten und Grenzen jüngerer Kinder bei der Herstellung einer Inklusionsrelation zwischen Begriffen. *Zeitschrift für Psychologie*, 197, 31-48.
- Goede, K. & Friedrich, M. (im Druck). Gefäße und Behälter: Das Phänomen der Exklusion bei Erwachsenen. *Zeitschrift für Psychologie*.
- Goede, K. & Friedrich, M. (1995). Wenn Rosen keine Blumen sind - Phänomene der Begriffsentwicklung. *Zeitschrift für Psychologie*, 203, 197-220.
- Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, 11, 23-63.
- Grossberg, S. (Ed.) (1982). *Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control*. Boston: Reidel Press.
- Grossberg, S. (1980). How does the brain build a cognitive code? *Psychological Review*, 87(1), 1-51.
- Grossberg, S. & Levine, D. (1975). Some developmental and attentional biases in the contrast enhancement and short term memory of recurrent neural networks. *Journal of Theoretical Biology*, 53, 341-380.
- Nelson, K., L. Rescorla, J. Gruendel & H. Benedict. (1978). Early lexicons: What do they mean? *Child Development*, 49, 960-968.
- Rosch, E.; Mervis, C.B.; Gray, W.D.; Johnson, D.M. & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.