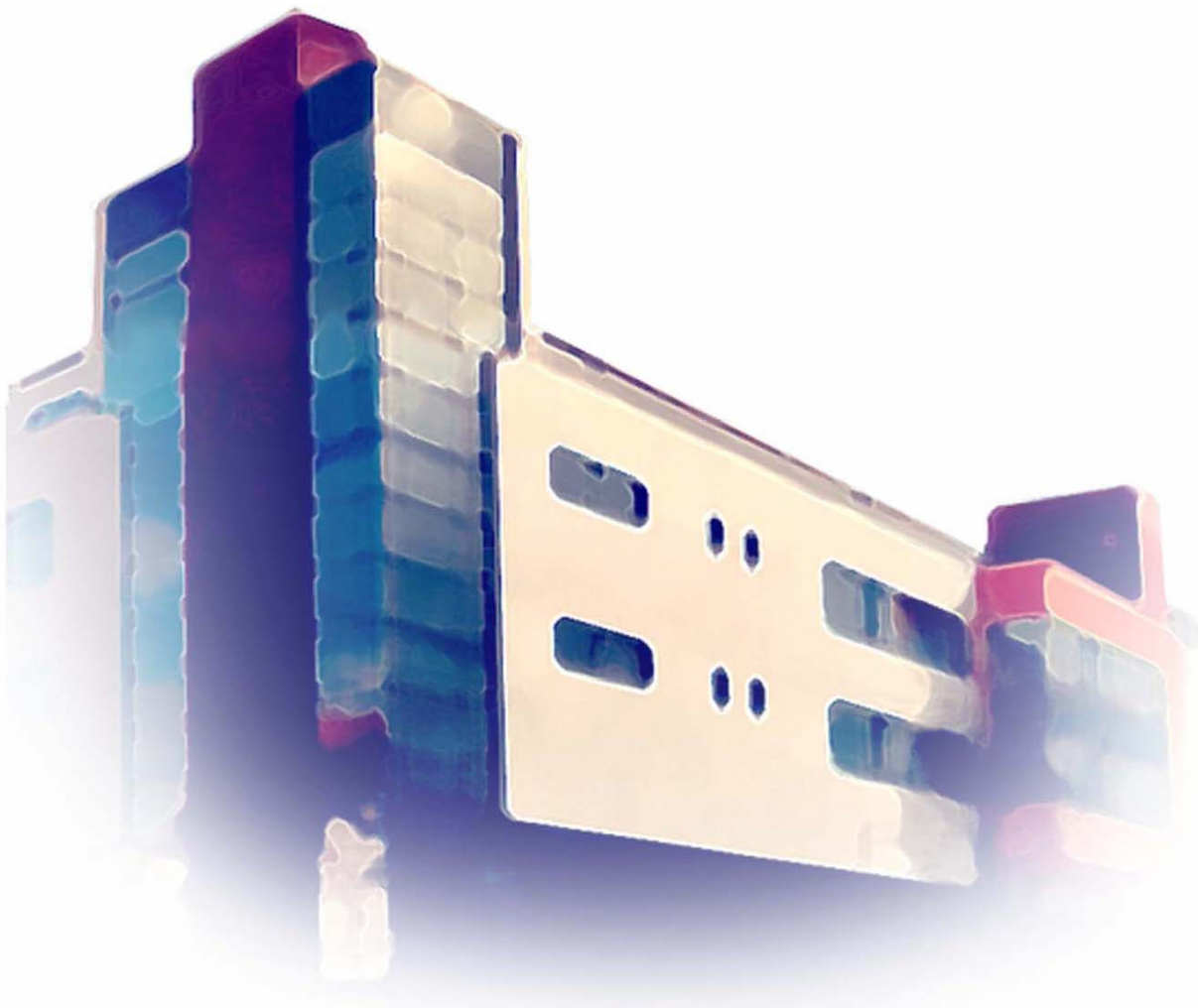


Roland Thomaschke

---

*Robert Cummins' Kritik an der kausalen Theorie  
mentaler Repräsentation am Beispiel der Lernbarkeit von  
LOT-Termen*

---



**PICS**

*Publications of the Institute of Cognitive Science*

*Volume 5-2004*

ISSN: 1610-5389

Series title: PICS  
Publications of the Institute of Cognitive Science

Volume: 5-2004

Place of publication: Osnabrück, Germany

Date: September 2004

Editors: Kai-Uwe Kühnberger  
Peter König  
Petra Ludewig

Cover design: Thorsten Hinrichs

**„Robert Cummins’ Kritik an der kausalen Theorie mentaler  
Repräsentation am Beispiel der Lernbarkeit von LOT-Termen“**

**STUDIENARBEIT**

zur Bachelorprüfung im Studiengang Cognitive Science

an der  
Universität Osnabrück

**Erster Gutachter: Prof. Dr. Achim Stephan**

**Zweiter Gutachter: Prof. Dr. Kai-Uwe Kühnberger**

**Verfasser: Roland Thomaschke**

Osnabrück, den 16. Feb. 2004



## Inhaltsverzeichnis

I. Einleitung	1
<b><u>1.0. Ziel der Arbeit</u></b>	1
<b><u>1.1. Theorien mentaler Repräsentation</u></b>	1
1.1.1. Mentale Repräsentation	1
1.1.2. Kausale Theorie mentaler Repräsentation	2
1.1.3. Cummins' Theorie mentaler Repräsentation	2
1.1.4. Die „Language of Thought“-Hypothese	3
<b><u>1.2. Cummins' Kritik an der kausalen Theorie mentaler Repräsentation</u></b>	3
1.2.1. Kausale Theorie und LOTH nach Cummins	4
1.2.2. Cummins' CT-Kritik	5
1.2.3. Cummins' LOTH-Kritik	6
1.2.4. Cummins' Argument zur Lernbarkeit von LOT-Termen	7
II. Lernbarkeit von LOT-Ausdrücken	7
<b><u>2.1. Die Texte</u></b>	7
2.1.1. Cummins (1997): <i>The LOT of the Causal Theory of Mental Content.</i>	7
2.1.2. Rupert, R. D. (2001): <i>Coining Terms in the Language of Thought: Innateness, Emergence, and the LOT of Cummins' s Argument against the Causal Theory of mental Content.</i>	8
2.1.3. Untersuchungsmethode	8
<b><u>2.2. Definitionen und Begriffe</u></b>	9
2.2.1. Transduktion von Eigenschaftsinstanzierungen	9
2.2.2. Definition der LOTH	10
2.2.3. Definition der CT	10
2.2.4. Implikation aus CT und LOTH	10
<b><u>2.3. Gelernte und native LOT-Ausdrücke</u></b>	11
2.3.1. Cummins' erster Argumentationsschritt	11
2.3.2. Kommentar	12
2.3.3. Cummins' erster Argumentationsschritt bei Rupert	19
2.3.4. Kommentar	19
2.3.5. Ruperts Kritik an Cummins' erstem Argumentationsschritt	20
2.3.6. Kommentar	26
<b><u>2.4. Lernen von LOT-Ausdrücken</u></b>	34
2.4.1. Cummins' zweiter Argumentationsschritt	34
2.4.2. Kommentar	35
2.4.3. Ruperts „coining terms“	35
2.4.4. Kommentar	38
III. Fazit	49
Literaturverzeichnis	53

# **I. Einleitung**

## **1.0. Ziel der Arbeit**

In der vorliegenden Arbeit werde ich die Kritik des Philosophen Robert Cummins an der kausalen Theorie mentaler Repräsentation, insbesondere an der „Language of Thought“-Hypothese, anhand von zwei Beispieltextrn exemplarisch darstellen und analysieren.

In Kapitel I. werde ich die Bedeutung der Texte im Rahmen der Philosophie der mentalen Repräsentation im Allgemeinen und in Cummins` s Kritik an der kausalen Theorie mentaler Repräsentation im Besonderen darstellen.

In Kapitel II. werde ich dann nach der Einführung einiger relevanter Grundbegriffe und der kurzen Vorstellung der beiden Texte diese nach Teilargumenten geordnet gegenüberstellen und die vorgebrachten Argumente kritisch überprüfen.

In Kapitel III. werde ich schließlich meine Ergebnisse aus dem zweiten Kapitel zusammenfassen und noch einmal in Bezug auf den im ersten Kapitel bereitgestellten Rahmen bewerten.

## **1.1. Theorien mentaler Repräsentation**

### **1.1.1. Mentale Repräsentationen**

Ein in der Philosophie des Geistes viel diskutiertes Problem ist die Intentionalität mentaler Prozesse und Zustände. Für Lösungsversuche dieses Problems werden häufig mentale Repräsentationen herangezogen. Dabei werden mentale Repräsentationen als Bestandteile mentaler Prozesse oder Zuständen, z. B. Gedanken, betrachtet, die ihrerseits auf eine Proposition, einen Gegenstand oder ein anderes Objekt gerichtet sind.

Obgleich zwar sowohl die Existenz mentaler Repräsentationen als auch deren Nützlichkeit für die Erklärung kognitiven Verhaltens in der gegenwärtigen Philosophie der Kognition nicht ganz unumstritten sind, stellen Repräsentationen einen substantiellen Bestandteil der überwiegenden Mehrheit wissenschaftlicher Theorien zur Erklärung kognitiver Systeme, die heute in der Philosophie der Kognition diskutiert werden, dar.

Dieser Benutzung des Repräsentationsbegriffes in der Philosophie der Kognition liegt keine einheitliche philosophische Theorie der Repräsentation zugrunde. Im Gegenteil konkurrieren in der Philosophie der Kognition viele unterschiedliche Ansätze um die Erklärung mentaler Repräsentation.

Als besonders schwierig scheint sich dabei die Frage herausgestellt zu haben, wie die Beziehung zwischen einer mentalen Repräsentation und ihrem Gehalt zu beschreiben ist.

So ist die Beantwortung dieser Frage für Theorien mentaler Repräsentation als wesentlich zu betrachten, weshalb solche Theorien meist danach klassifiziert werden, wie sie die Beziehung zwischen Repräsentation und Gehalt erklären.

Im Folgenden geht es nun darum, ob eine bestimmte Klasse von Theorien, nämlich die kausalen Theorien mentaler Repräsentation, aus prinzipiellen Gründen als unzulänglich zu betrachten sind. In Kapitel 1.1.2. werde ich deshalb kurz darstellen, was kausale Theorien charakterisiert.

Vorgebracht wird die hier zu untersuchende Kritik an der kausalen Theorie mentaler Repräsentation von Robert Cummins, welcher folglich eine Theorie mentaler Repräsentation ver-

tritt, die nicht als kausal zu klassifizieren ist. Diese Theorie werde ich in 1.1.3. kurz skizzieren um dann in 1.1.4. eine Hypothese bezüglich der Struktur mentaler Repräsentationsschemen, welche in Cummins' Kritik eine entscheidende Rolle spielt, nämlich die „Language of Thought“- Hypothese, einzuführen.

### **1.1.2. Kausale Theorie mentaler Repräsentation**

Kausale Theorien mentaler Repräsentation sind dadurch gekennzeichnet, dass sie die Beziehung zwischen Repräsentation und Gehalt als Kausalrelation beschreiben.

So kann jeder mentalen Repräsentation genau der Gehalt zugeordnet werden, der zu dieser Repräsentation in einer genau spezifizierten Verursachungsbeziehung steht. Es werden verschiedene Versionen der kausalen Theorie vertreten, welche sich dahingehend unterscheiden, wie genau die kausale Beziehung zu beschreiben ist, die einer Repräsentation ihren Gehalt zuordnet. Viel diskutierte Vertreter der kausalen Theorie mentaler Repräsentation sind z.B. Dretske (1981) und Fodor (1987).

Cummins kritisiert in dem hier zu untersuchenden Text jedoch nicht eine bestimmte kausale Theorie, sondern versucht die kausale Theorie in ihrer allgemeinsten Form zu widerlegen. Eine Explikation dieser allgemeinen Version der kausalen Theorie ist in Kapitel 2.2.3. zu finden.

Im Folgenden werde ich die kausale Theorie mentaler Repräsentation, Cummins (1997) und Rupert (2001) folgend, abkürzend als CT<sup>1</sup> bezeichnen.

### **1.1.3. Cummins' Theorie mentaler Repräsentation**

Obwohl Cummins in Cummins (1997), von seinem Vorgehen in anderen Kommentaren zur CT abweichend, darauf verzichtet, der kritisierten Theorie Teile seiner eigenen Theorie mentaler Repräsentation gegenüberzustellen, werde ich seine Theorie hier trotzdem kurz vorstellen.

Cummins' Theorie mentaler Repräsentation ist bezüglich des oben genannten Klassifikationskriteriums für solche Theorien – Beschreibungsweise der Beziehung zwischen Repräsentation und Gehalt – der Klasse der sogenannten Bildtheorien zuzuordnen.

Bildtheorien beschreiben die Beziehung zwischen Repräsentation und Gehalt als eine bestimmte Art von Ähnlichkeit. Solche Theorien werden heute seltener vertreten als kausale Theorien. Andere Vertreter von Bildtheorien sind z.B. Kosslyn (1980) oder McGinn (1989).

Cummins unterteilt den intentionalen Gehalt einer Repräsentation in zwei Teil-Gehalte.

---

<sup>1</sup> Cummins und Rupert sprechen von Theorien mentalen Gehaltes anstatt von Theorien mentaler Repräsentation. So fungiert der Ausdruck CT in den hier zu untersuchenden Texten von Cummins und Rupert auch als Abkürzung für „causal theory of mental content“. Die Bezeichnung ist bei Cummins wahrscheinlich als eine präzisere Spezifikation des Explanandums der kausalen Theorie zu verstehen. Die kausale Relation erklärt schließlich in einer vollständigen Theorie mentaler Repräsentation nur die Zuordnung des Gehaltes zur entsprechenden Repräsentation, aber nicht, was z.B. der Träger einer Repräsentation ist oder wie Repräsentationen in Erklärungen von Gedanken o.ä. einfließen.

Ich habe hier Cummins anderen Texten (Cummins: 1996a, 1996b, Cummins; Poirier: forthcoming) folgend, den Ausdruck „Theorie mentaler Repräsentation“ gewählt, da der Bezug auf mentalen Gehalt auch irreführend als erweiternd verstanden werden kann. Cummins hält nämlich eine kausale Erklärung für den Gehalt anderer mentaler – nicht repräsentierender – Entitäten durchaus für angemessen. Ausdrücklich wird dies in Cummins; Poirier (forthcoming) für Indikatoren betont.

Einer in diesem Sinne angemessenen Eingrenzung des Geltungsbereichs der hier zu untersuchenden kausalen Theorie würde somit eher die Bezeichnung „causal theory of content of mental representations“ gerecht werden. Jedenfalls bezieht sich mein Ausdruck „kausale Theorie mentaler Repräsentation“ in diesem Kontext, genauso wie Cummins' „causal theory of mental content“, auf eine Theorie über den Gehalt mentaler Repräsentationen und wird deshalb gleichfalls mit CT abgekürzt.

Einerseits ist eine Repräsentation auf ein Ziel („target“) gerichtet. Die Relation zwischen Repräsentation und Ziel wird durch die Funktion der Repräsentation bestimmt. Wenn die Repräsentation in einem kognitiven System die Funktion hat einen bestimmten Gegenstand zu repräsentieren, ist dieser Gegenstand das Ziel der Repräsentation und bestimmt damit einen Teil des Gehaltes.

Andererseits ist eine Repräsentation auf einen Inhalt („content“) gerichtet. Die Relation zwischen Inhalt und Repräsentation wird als Isomorphismus zwischen den Strukturen von Inhalt und Repräsentation beschrieben. Ist die Struktur einer Repräsentation zur Struktur eines bestimmten Objektes isomorph, dann ist dieses Objekt der Inhalt der Repräsentation und bestimmt damit den anderen Teil des Gehaltes.

Diese Theorie ist von Cummins in Cummins (1996a) ausführlich dargestellt und in Cummins (2000, 2002, forthcoming) gegen Einwände verteidigt und erweitert worden.

Da es in dem hier relevanten Text (Cummins 1997) jedoch ausschließlich um die CT geht, werde ich Cummins' Theorie hier nicht weiter ausführen.

#### **1.1.4. Die „Language of Thought“- Hypothese**

Die „Language of Thought“- Hypothese geht auf Fodor (1975) zurück und besteht in einer Behauptung bezüglich der Struktur eines Schemas mentaler Repräsentationen.

Im Folgenden werde ich mich auf die „Language of Thought“- Hypothese mit der Abkürzung LOTH beziehen<sup>2</sup>.

Gilt für ein Schema mentaler Repräsentationen die LOTH, dann sind die Repräsentationen in diesem Schema gemäß einer sprachähnlichen Syntax angeordnet und besitzen eine kompositionelle Semantik. Das bedeutet, dass es eine Menge von atomaren Grundtermen gibt, aus denen gemäß der Syntax komplexe Terme gebildet werden können, und dass der Gehalt einer komplexen Repräsentation ihrer Syntax entsprechend aus den Gehalten der die Repräsentation konstituierenden Grundterme zusammengesetzt ist.

Über die Art der Beziehung zwischen Repräsentation und Gehalt sagt die LOTH nichts aus.

Eine genaue Definition der LOTH- Version, welche Cummins in Cummins (1997) kritisiert, ist in Kapitel 2.2.2. enthalten.

### **1.2. Cummins' Kritik an der kausalen Theorie mentaler Repräsentation**

Cummins rechtfertigt und motiviert seine eigene Theorie mentaler Repräsentation zum großen Teil durch Kritik an anderen Theorien der gegenwärtigen Philosophie mentaler Repräsentation. Alle gegenwärtigen Theorien, außer seiner eigenen, sind Cummins zufolge aus prinzipiellen Gründen unzulänglich. Cummins begründet dies so, dass alle gegenwärtigen Theorien repräsentationalen Gehalt durch den richtigen Gebrauch der entsprechenden Repräsentation definieren, das heißt „use theories“ sind.

„I think *all* of the theories currently taken seriously are use theories of one kind or another.“ (Cummins 1996a, S. 29). „Use theories“ lassen, Cummins' Ansicht zufolge, keinen repräsentationalen Fehler zu. Diesen sieht er jedoch als für das Lernen von mentalen Repräsentationen unerlässlich an. Somit können „use theories“ das Zustandekommen von Repräsentationen nicht vollständig erklären.

Cummins übt allerdings auch eine Menge detailliertere Kritik an den spezifischen Theorien anderer Philosophen oder an enger eingegrenzten Theoriegruppen.

---

<sup>2</sup> In der deutschsprachigen Literatur wird „Language of Thought“ unterschiedlich übersetzt: mit z.B. „Sprache des Denkens“ (Carrier, M. 1996) oder „Sprache des Geistes“ (Beckermann, A. 2001). Da die Abkürzung LOTH, manchmal auch LOT (siehe Fußnote 5), in der gegenwärtigen Philosophie der Kognition sehr verbreitet ist, habe ich hier auf eine Übersetzung verzichtet.



Dabei nimmt die Kritik an der kausalen Theorie, da diese in der gegenwärtigen Philosophie mentaler Repräsentation weit verbreitet ist (siehe 1.1.2.), einen erheblichen Raum ein.

In den Kapiteln 1.2.1. bis 1.2.3. werde ich einen kurzen Überblick über Cummins verschiedene Ansätze zur Kritik der CT geben.

Eine wichtige Rolle spielt in Cummins' CT-Kritik die LOTH, da diese Cummins zufolge von jeder CT impliziert wird. Cummins Argumente für diese Implikationsbeziehung werde ich in 1.2.1 skizzieren.

So ist jede Kritik an der LOTH bei Cummins auch als Kritik an der CT zu verstehen. LOTH Kritik, welche nicht speziell Bezug auf Annahmen der CT nimmt, werde ich in 1.2.3. nennen.

In 1.2.2. werde ich die verschiedenen Ansätze referieren, mit denen Cummins die CT explizit kritisiert und in 1.2.4. die im Hauptteil besprochenen Texte in Cummins' CT-Kritik einordnen.

### **1.2.1. Kausale Theorie und LOTH nach Cummins**

Cummins' Argumente für eine Implikationsbeziehung zwischen CT und LOTH sind zuerst in Cummins (1996a) und noch einmal erweitert in Cummins (1997) zu finden.

Da Cummins behauptet, dass jede CT die Grundeigenschaften der LOTH impliziert, wird für die Argumentation von der CT nur vorausgesetzt, dass der Gehalt eines repräsentationalen Grundterms<sup>3</sup> durch Detektion fixiert ist. Daraus folgt nach Cummins, dass die Repräsentationen „syntactically individuated“ und von endlicher Anzahl sind, sowie dass das repräsentationale Schema eine kompositionelle Semantik besitzen muss. Die Bedingungen, gemäß einer Syntax angeordnet zu sein und eine kompositionelle Semantik zu haben, können als Kerneigenschaften der LOTH für mentale Repräsentationen bezeichnet werden (siehe 1.1.4. und 2.2.2.).

Dass die Repräsentationen unter Annahme der CT beliebig sind, begründet Cummins in Cummins (1996a, S. 69) und Cummins (1997, S. 539) damit, dass das Funktionieren einer Detektion<sup>4</sup> nicht von der Struktur des Ausgabe-Symbols abhängig ist.

In Cummins (1996a, S. 69f.) argumentiert Cummins direkt für die Implikation der kompositionalen Semantik durch die CT und lässt offen, ob die Menge der repräsentationalen Grundterme endlich ist:

Nach Cummins ist es nicht möglich, dass der Gehalt komplexer Repräsentationen so, wie der Gehalt der Grundterme, durch Detektion bestimmt wird.

Denn wäre der Gehalt komplexer Symbole auch durch Detektion bestimmt, könnte für jede komplexe Repräsentation der Gehalt ihrer Konstituenten, wegen der Beliebigkeit der Detektor-Symbole, auch völlig unabhängig von dem Gehalt der komplexen Repräsentation sein, was nach Cummins (1996a, S. 69 f.) im Widerspruch zur Komplexität des Symbols stände. Da die komplexen Repräsentationen ihren Gehalt nicht aus Detektion bekommen können, muss der Gehalt der komplexen Symbole aus den Gehalten der Grundterme gemäß ihres syntaktischen Zusammenhangs zusammengesetzt sein. Dies ist nach Cummins gleichbedeutend mit kompositioneller Semantik.

---

<sup>3</sup> Der Ausdruck „repräsentationaler Grundterm“ ist in 1.1. bisher nur in der Beschreibung der LOTH aufgetaucht. „Repräsentationaler Grundterm“ bedeutet hier nur, dass die Repräsentation nicht semantisch aus anderen zusammengesetzt ist, ohne dass damit schon eine kompositionelle Semantik, und damit teilweise die LOTH, vorausgesetzt würde. Bei Cummins und anderen wird ein repräsentationaler Grundterm als „representational primitive“ bezeichnet. Wird die Geltung der LOTH vorausgesetzt, bezeichne ich einen repräsentationalen Grundterm als LOT-Grundterm. Siehe zu dieser Bezeichnung auch Fußnote 5.

<sup>4</sup> Mit Detektion bezeichnet Cummins (1996a, 1997, 2000, forthcoming) die korrekte Instanzierung („tokening“) einer Repräsentation bei Geltung der CT. Im Hauptkapitel werde ich Cummins in dieser Bezeichnungsweise folgen.

In Cummins (1997) wird zuerst für die Endlichkeit der Menge der repräsentationalen Grundterme argumentiert und daraus die Notwendigkeit einer kompositionellen Semantik abgeleitet. Die Endlichkeit der Menge der repräsentationalen Grundterme begründet Cummins damit, dass jeder Grundterm von einem Detektor detektiert werden müsse. Weder könne jedoch ein Detektor unendlich viele Grundterme detektieren noch könnten im menschlichen Gehirn, aufgrund der Begrenzung des Speicherplatzes, unendlich viele Detektoren vorhanden sein. Also müsse die Menge der repräsentationalen Grundterme endlich sein.

In Cummins (1997, S. 540f) schließt Cummins von der Endlichkeit der Menge der repräsentationalen Grundterme und der Möglichkeit, eine unendliche Menge von semantisch unterschiedenen Repräsentationen zu bilden, dann direkt auf die Notwendigkeit einer kompositionellen Semantik.

Daraus, dass CT eine endliche Menge von semantisch beliebigen Grundtermen impliziert, welche mittels kompositionaler Semantik unendlich viele semantisch unterschiedliche Terme hervorbringen kann, folgert Cummins, dass CT die LOTH impliziert.

„A scheme of finitely many semantically arbitrary primitives which allows for an unbounded set of semantically distinct complex representations whose contents are determined by their constituents and mode of combinations LOT as near as makes no difference. So CT entails LOT”<sup>5</sup> (Cummins 1997, S. 541).

### **1.2.2. Cummins' CT-Kritik**

Bei Cummins finden sich drei unterschiedliche Ansätze zur Kritik der CT.

Erstens versucht er in Cummins (1996a, S. 53 – 62) nachzuweisen, dass CTs grundsätzlich „use theories“ sind.

CTs bestimmen den repräsentationalen Gehalt durch Bestimmung einer kausalen Relation zwischen Repräsentation und Gehalt (siehe 1.1.2.). Eine solche Bestimmung kann nach Cummins als Bestimmung einer korrekten Benutzung der Repräsentation aufgefasst werden. Also sind CTs nach Cummins „use theories“.

Da er „use theories“ aus oben genannten Gründen für prinzipiell unzulänglich hält, ist die CT für Cummins durch den Nachweis, eine „use theorie“ zu sein, widerlegt.

Zweitens stellt Cummins eine „explanatorische Bedingung“ („explanatory constraint“ Cummins 1996a, S. 2) für eine Theorie mentaler Repräsentation auf, von welcher er in Cummins (1996a; 2000; forthcoming ) nachweist, dass die CT dieser Bedingung nicht gerecht werden kann. Cummins ist der Auffassung, wenn eine Theorie mentaler Repräsentation, statt nur zur Beschreibung, wirklich zur Erklärung kognitiver Prozesse beitragen sollte, dann müsse diese Theorie mentalen Repräsentationen intrinsische Informationen über das Repräsentierte zuschreiben.

Die CT kann diese Bedingung nach Cummins nicht erfüllen. Da für die Möglichkeit einer kausalen Relation zwischen Repräsentation und Gehalt die physische Struktur der Repräsentation keine wesentliche Rolle spielt, ist das Symbol für einen bestimmten repräsentationalen Gehalt beliebig, kann also keine intrinsischen strukturellen Merkmale haben, welche Informationen über den Gehalt tragen könnten. Somit ist das, was die CT als Repräsentation auffasst,

---

<sup>5</sup>Cummins benutzt die Abkürzung LOT sowohl für „Language of Thought“ selbst, als auch – wie in der zitierten Passage – für die LOTH, also die These, dass mentale Repräsentationen in einer „Language of Thought“ angeordnet sind. Ich beziehe mich im Folgenden mit der Abkürzung LOT ausschließlich auf die „Language of Thought“ selbst, anstatt auf die LOTH. Falls Cummins oder Rupert in hier zitierten Passagen die Abkürzung LOT verwenden, um auf die LOTH zu verweisen, werde ich jeweils darauf hinweisen.

nach Cummins bloß ein nichtrepräsentierendes Detektor-Symbol, da die explanatorische Bedingung nicht erfüllt ist.

Drittens kritisiert Cummins die CT in Cummins (1997) im Zusammenhang mit der LOTH. Dort versucht er zu zeigen, dass die LOTH, welche ja Cummins zufolge von der CT impliziert wird, unter Annahme der CT im Widerspruch mit elementaren Erkenntnissen der Psychologie steht. Dieser Ansatz wird in den beiden Hauptkapiteln besprochen.

### **1.2.3. Cummins' LOTH-Kritik**

Die LOTH wird von Cummins in den Aufsätzen Cummins (1996b), Cummins et al. (2001) und Cummins et al. (forthcoming) auch ohne Bezug auf die CT kritisiert. Dies erfolgt allerdings nur indirekt durch Kritik an einem wesentlichen Argument für die LOTH, kommt hier aber trotzdem zur Sprache, da ich in beiden Hauptkapiteln darauf zurückgreife.

Das Argument ist das sogenannte Systematizitätsargument. Die Version dieses Arguments, auf welche sich Cummins in ebengenannten Texten bezieht, war von Fodor und Pylyshyn in Fodor et al. (1988) vorgebracht worden und ist dann von Fodor und McLaughlin in Fodor et al. (1990) bezüglich der Kritik von (unter anderen) Smolensky (1990) modifiziert worden. Das Systematizitätsargument schließt von unserer Fähigkeit Sprachen zu verstehen, welche Systematizität<sup>6</sup> besitzen, auf die Systematizität von Gedanken<sup>7</sup> und weiter darauf, dass mentale Repräsentationen in einem der natürlichen Sprache ähnlichen syntaktischen Schema angeordnet sind und kompositionelle Semantik haben.

Die Ergebnisse von Cummins' kritischer Beschäftigung mit dem Systematizitätsargument lassen sich vereinfacht so zusammenfassen:

Falls das Systematizitätsargument für natürliche Sprache gilt, muss es auch für andere Bereiche gelten, die so kognitiv verarbeitet werden können, dass man auf eine Systematizität der Gedanken bezüglich dieser Bereiche schließen kann, z.B. dreidimensionale Bilder.

Da die Struktur der Rekombinierbarkeit der Konstituenten von komplexen Ausdrücken in den verschiedenen Bereichen nicht isomorph ist, würde das Systematizitätsargument also ergeben, dass Gedanken bezüglich verschiedener nicht isomorpher struktureller Schemata Systematizität besitzen würden<sup>8</sup>. Um dies gemäß des Systematizitätsarguments zu erklären, müsste man mehrere repräsentationale Schemata, die nicht aufeinander reduziert werden könnten (da ja die systematizitätsrelevanten Strukturen der kognizierten Bereiche nicht isomorph sind), annehmen.

Dies würde bedeuten, dass sich mentale Repräsentationen nicht als in einem einzigen strukturellen Schema mit sprachähnlicher Syntax angeordnet beschreiben ließen.

Somit kann das Systematizitätsargument nach Cummins nicht als Argument für die LOTH benutzt werden.

---

<sup>6</sup> In der hier skizzierten Diskussion tauchen unterschiedliche Definitionen von Systematizität auf. Allgemein ist Systematizität eine Eigenschaft eines Systems von Ausdrücken (Sprache, repräsentationales Schema), die dann vorliegt, wenn das System mit jedem seiner Ausdrücke auch dessen durch grammatisch korrekte Rekombination entstandene Varianten enthält.

Dass natürliche Sprachen Systematizität besitzen, wird in der gesamten hier angesprochenen Diskussion, insbesondere in den Texten von Cummins, vorausgesetzt. Erst in den letzten Jahren ist diese Voraussetzung mehrfach angezweifelt worden, z.B. in Johnson (forthcoming).

<sup>7</sup> Auf die Systematizität von Gedanken schließt das Systematizitätsargument, wenn folgende Situation vorliegt: „[...] whenever it [das kognitive System] can process a sentence *s*, it can process systematic variants of *s*, where systematic variation is understood in terms of permuting constituents or (more strongly) substituting constituents of the same grammatical category.“ (Cummins 1996b, S. 594).

<sup>8</sup> Dieser Schluss gilt natürlich nur, falls das kognitive System die Bereiche entsprechend dem in Fußnote 7 genannten Kriterium verarbeitet.

#### **1.2.4. Cummins' Argument zur Lernbarkeit von LOT-Termen**

Zur genaueren Untersuchung von Cummins CT-Kritik werde ich die Aufsätze Cummins (1997): *The LOT of the Causal Theory of Mental Content* und Rupert (2001): *Coining Terms in the Language of Thought: Innateness, Emergence, and the LOT of Cummins' s Argument against the Causal Theory of mental Content* analysieren.

Aus mehreren Gründen eignen sich diese Texte besonders gut für eine Untersuchung von Cummins' CT-Kritik.

Erstens handelt es sich bei dem Argument, welches Cummins in Cummins (1997) vorbringt nicht, wie z.B. in Cummins (1996b) oder Cummins; et al. (2001), um einen Beitrag zu einer in der Philosophie mentaler Repräsentation schon länger bestehenden Debatte, sondern um eine Möglichkeit der Kritik an der CT, welche vor Cummins (1997) noch nicht formuliert worden ist.

Zweitens ist das in Cummins (1997) vorgestellte Argument für Cummins' gesamte CT-Kritik insofern charakteristisch, als dass hier mit Bezug auf die Möglichkeit von Lernvorgängen argumentiert wird.

Drittens antwortet Rupert Cummins in Rupert (2001) direkt und differenziert auf sein in Cummins (1997) vorgebrachtes allgemeines Argument gegen die CT, ohne eine vollständige, unabhängige kausale Theorie mentaler Repräsentation zu entwickeln. Aufgrund der Einheitlichkeit der Abstraktionsebene eignen sich die Texte gut als Ausgangspunkt für eine eigene Beurteilung von Cummins' Argumentation.

Viertens ist eine Untersuchung der beiden Texte deshalb sinnvoll, weil diese Debatte zwischen Rupert und Cummins bisher in der Fachliteratur nicht kommentiert worden ist.

## **II. Lernbarkeit von LOT-Ausdrücken**

### **2.1. Die Texte**

Die beiden hier zu untersuchenden Texte sind beide im *Journal of Philosophy* erschienen. Während sich der 1997 erschienene Text von Cummins nicht speziell auf andere Publikationen bezieht, sondern die CT allgemein kritisiert, stellt der 2001 erschienene Text von Rupert eine direkte Kritik an Cummins' Aufsatz dar.

Ich werde in diesem Kapitel beide Texte kurz vorstellen und die bei der Untersuchung derselben angewandte Methode erläutern.

#### **2.1.1 Cummins (1997): *The LOT of the Causal Theory of Mental Content***

In dem Text: *The LOT of the Causal Theory of Mental Content* von 1997 konstruiert Cummins ein Argument gegen die kausale Theorie mentaler Repräsentation. Vorausgesetzt wird die Unmöglichkeit der direkten Transduktion von Eigenschaftsinstanzierungen in der Außenwelt als psychologische Tatsache. Cummins zufolge ist die CT abzulehnen, da sie mit der eben genannten Tatsache im Widerspruch steht.

Diesen Widerspruch konstruiert Cummins in zwei Teilargumenten.

Zuerst zeigt er, dass die CT nicht mit der LOT-Hypothese und der Unmöglichkeit der direkten Transduktion von Eigenschaftsinstanzierungen gemeinsam wahr sein kann.

Dies wird damit begründet, dass mentale Repräsentationen CT zufolge durch Detektion der zu repräsentierenden Eigenschaften erzeugt werden, hierfür aufgrund der Unmöglichkeit der direkten Transduktion von Eigenschaftsinstanzierungen aber eine Theorie notwendig ist. Solche Theorien müssen nach Cummins oft gelernt werden, was LOTH zufolge bedeutet, dass die

entsprechende Theorie als Menge von LOT-Sätzen, welche die betreffende Repräsentation enthalten, formuliert und dann durch Erfahrung bestätigt wird.

Für die zu formulierende Theorie müsste die zu lernende Repräsentation allerdings schon erzeugt werden können, was ja durch die Formulierung der Theorie erst gelernt werden soll.

So ergibt sich aus den drei Voraussetzungen, der CT, der LOTH und der Unmöglichkeit der direkten Transduktion von Eigenschaftsinstanzierungen, ein Widerspruch.

Das zweite Teilargument besteht darin, zu zeigen, dass die kausale Theorie mentaler Repräsentationen die LOTH impliziert. Diese Argumentation ist in Kapitel 1.2.1 bereits dargestellt worden.

Für diese beiden Teilargumente nennt Cummins noch einige potentielle Gegenargumente und versucht, diese zu entkräften.

Wenn man Cummins folgend annimmt, dass erstens die Unmöglichkeit der direkten Transduktion von Eigenschaftsinstanzierungen eine wahre Tatsache ist und zweitens diese Tatsache nicht zusammen mit der CT und der LOTH wahr sein kann und drittens dass die CT die LOTH impliziert, dann folgt logisch, dass die CT falsch ist.

### **2.1.2 Rupert, R.D.(2001): *Coining Terms in the Language of Thought: Innateness, Emergence, and the LOT of Cummins` s Argument against the Causal Theory of mental Content.***

Ruperts Text *Coining Terms in the Language of Thought: Innateness, Emergence, and the LOT of Cummins` s Argument against the Causal Theory of mental Content* (2001) ist eine Verteidigung der CT, welche sich direkt gegen Cummins` eben genanntes Argument richtet. Dabei konzentriert sich Rupert auf die Widerlegung des erste Teilargumentes. Er argumentiert also dafür, dass die CT mit der LOTH und der Unmöglichkeit der direkten Transduktion von Eigenschaftsinstanzierungen vereinbar ist.

Zuerst versucht Rupert darzulegen, dass überhaupt kaum LOT-Terme gibt, deren korrekte Instanzierung eine Theorie erfordert, welche gelernt werden muss.

Des weiteren skizziert Rupert einen Weg, auf dem solche Theorien gelernt werden können, ohne, dass dies schon die korrekte Detektion der entsprechenden Repräsentation voraussetzen würde. Damit wäre der von Cummins konstruierte Widerspruch zwischen CT, LOTH und der Unmöglichkeit der direkten Transduktion von Eigenschaftsinstanzierungen aufgelöst.

### **2.1.3 Untersuchungsmethode**

Cummins` erstes Teilargument, um welches es hier gehen soll, lässt sich in zwei Argumentationsschritte unterteilen, die von Rupert in seinem Verteidigung der CT auch getrennt kritisiert werden.

Um Cummins` Argument direkter mit Ruperts Kritik zu konfrontieren, werde ich die Texte nicht nacheinander diskutieren, sondern Cummins` und Ruperts Beiträge bezüglich der beiden Argumentationsschritte jeweils direkt gegeneinander stellen und dann auch gleich bezüglich der einzelnen Argumente kommentieren.

In 2.3 erläutere und kommentiere ich Cummins` erste Argumentationsschritt und Ruperts Kritik daran.

Das selbe geschieht mit dem zweiten Argumentationsschritt im Kapitel 2.4..

## 2.2. Definitionen und Begriffe

Im Folgenden werde ich einige Begriffe einführen, welche bei Cummins und Rupert explizit definiert sind und deren Gebrauch in beiden Texten einheitlich ist. Dies sind eine von beiden Autoren als für die Diskussion maßgeblich anerkannte Formulierung der LOTH und eine allgemeine Charakterisierung der Kerneigenschaft der CT. Des Weiteren die Formulierung der Unmöglichkeit der direkten Transduktion von zu repräsentierenden Eigenschaftsinstanzierungen sowie eine von beiden akzeptierte Folgerung aus CT und LOTH.

### 2.2.1. Transduktion von Eigenschaftsinstanzierungen

Die Unmöglichkeit der direkten Transduktion zu repräsentierender Eigenschaftsinstanzierungen wird bei Cummins und Rupert als „nontransducibility of distal properties“ bezeichnet (Cummins 1997, S. 535; Rupert 2001, S. 501) und von mir, Rupert und Cummins folgend, im weiteren Text mit NTDP abgekürzt.

Für eine Definition von „Transduktion“ im hier relevanten Sinne verweist Cummins auf Pylyshyn (1984)<sup>9</sup>. Die Rolle der Transduktion in Pylyshyns Theorie der Kognition soll hier kurz skizziert werden. Pylyshyn unterscheidet drei Ebenen der Beschreibung kognitiver Systeme und Teilsysteme: die physikalische Ebene, die funktionale Ebene (auch symbolische) und die repräsentationale (auch semantische) Ebene. Nach Pylyshyn ist für die Beschreibung eines kognitiven Vorgangs die Ebene zu wählen, auf welcher sich die dem Vorgang zugrundeliegende Regularität am einfachsten ausdrücken lässt. Eine Transduktion ist ein Vorgang, der selber auf der physikalischen Ebene zu beschreiben ist, und der Ereignisse, welche ebenfalls auf der physikalischen Ebene zu beschreiben sind, auf Ereignisse, die auf der funktionalen Ebene zu beschreiben sind, abbildet. Ein paradigmatischer Vorgang ist eine retinale Photorezeption. Nach Cummins ist es eine psychologische Tatsache, dass die Detektion einer Eigenschaftsinstanzierung in der Umwelt nicht als ein solches Ereignis bezeichnet werden kann.

Eine Implikation der NTDP ist, Cummins zufolge, dass eine Person um eine instanziierte Eigenschaft zu detektieren, eine Theorie über diese Eigenschaft besitzen muss. Theorie bedeutet hier nicht, dass die Person explizit formulierte Sätze, welche Folgerungen und Vorhersagen ermöglichen, wissen muss, - dies ist unter anderem Gegenstand der folgenden Diskussion – sondern nur, dass die Person irgend eine Art Wissen über die zu detektierende Eigenschaft haben muss, welches es ihr ermöglicht eine Instanzierung der Eigenschaft korrekt festzustellen. „Wissen“ wird hier nicht im philosophischen Sinne als gerechtfertigte wahre Meinung gebraucht, sondern: „Knowledge in this sense is information that is treated as if it were reliable by the system that uses it.“ (Cummins 1997, S. 536, Fußnote 6).

Eine Theorie, welche die korrekte Detektion einer Eigenschaftsinstanzierung ermöglicht werde ich hier, Rupert und Cummins folgend, mit MT – für „mediating theory“ – abkürzen.

Wenn Cummins und Rupert mit NTDP argumentieren, beziehen sie sich in jedem Fall direkt auf die eben beschriebene Folgerung, dass zur Detektion einer Eigenschaft eine Theorie über die zu detektierende Eigenschaft notwendig ist. So führt Rupert NTDP gleich ein als „[...] that theory mediates sensory detection.“ (Rupert 2001, S. 499).

Deshalb werde ich mich im Folgenden mit NTDP, Cummins und Rupert folgend, auf die Aussage beziehen, dass eine Person, um eine Eigenschaftsinstanzierung zu detektieren, eine Theorie im eben beschriebenen Sinne über die betreffende Eigenschaft besitzen muss.

NTDP wird von Cummins als „an elementary fact of perceptual psychology“ bezeichnet und für seine Argumentation als unbezweifelbare Tatsache vorausgesetzt.

---

<sup>9</sup> Auf Pylyshyns Theorie der Kognition werde ich in 2.3.2 genauer eingehen.

In dieser Einschätzung folgt ihm Rupert ohne Einschränkung:

„Currently accepted theories of human perception imply that, for a subject reliably to detect members of a kind, she must possess a theory of that kind [...].“ (Rupert 2001, S. 500).

### **2.2.2 Definition der LOTH**

Cummins charakterisiert die LOTH in Cummins (1997) als die Behauptung, dass alle mentalen Repräsentationen des menschlichen Geistes als Teile eines repräsentationalen Schemas anzusehen sind, welches folgende drei Eigenschaften erfüllt:

(LOTH 1) Das Schema enthält eine endliche Anzahl semantisch atomarer Grundterme<sup>10</sup>, die syntaktisch unterschieden sind.

(LOTH2) Jeder Term ist eine Verkettung von atomaren Grundtermen.

(LOTH3) Der Gehalt jedes komplexen Terms ist der Wert einer Funktion, deren Argumente die Gehalte der Grundterme, aus denen der komplexe Term besteht, und die Syntax des Terms sind.

Für diese Definition beruft sich Cummins auf Fodor (1975). Rupert nimmt in Rupert (2001) an dieser Definition keine Ergänzungen oder Veränderungen vor.

### **2.2.3. Definition der CT**

Cummins definiert die CT in Cummins (1997) folgendermaßen:

„By CT, I mean the doctrine that the contents of the semantic primitives in the human scheme of mental representation are determined by their role in detection. The basic idea is just that the content of a primitive  $r$  in a system  $\Sigma$  is the property  $P$  if there is the right kind of causal connection between instantiations of  $P$  and tokenings of  $r$  by  $\Sigma$ 's detectors.“ (Cummins 1997, S. 535).

Rupert bezweifelt in Rupert (2001) zwar, dass die Eigenschaft, die Gehalte der repräsentationalen Grundterme durch die Rolle der Terme in einer Detektionsfunktion des repräsentierenden Systems (nämlich der Rolle, in einer theoriespezifischen kausalen Relation zur Instanzierung einer bestimmten Eigenschaft zu stehen) zu bestimmen, genau die Eigenschaft ist, welche die CT von anderen Theorien unterscheidet. Da diese Eigenschaft jedoch vielen bekannten CTs ohne Zweifel zukommt, akzeptiert Rupert die oben genannte Charakterisierung als Diskussionsgrundlage: „[...] the literature contains enough theories of this sort to merit careful and critical consideration of Cummin's argument.“ (Rupert 2001, S. 500).

### **2.2.4. Implikation aus CT und LOTH**

Im zweiten Argumentationsschritt in Cummins (1997) sowie in Cummins (1996a) argumentiert Cummins, dass die CT eine minimale Version der LOTH impliziert. Diese These wird von Rupert in Rupert (2001) nicht explizit besprochen. Für die Diskussion von Cummins' erstem Teilargument, um die es im Folgenden gehen wird, setzt Rupert also LOTH sowie CT voraus, unabhängig davon, ob die von Cummins behauptete Implikationsbeziehung gilt:

„While his [Cummins'] argument merits attention, I set it aside here. I do not defend CT by attempting to sever it from LOT, as causal theorists might propose to do in response to

---

<sup>10</sup> „semantisch atomare Grundterme“ bezeichnet Cummins mit „sematically primitive expressions“ (siehe auch Fußnote 3).

Cummins `s critique of CT; consequently, for present purposes, I grant LOT `s existence.”<sup>11</sup> (Rupert 2001, S. 500, Fußnote 2)

Eine Folgerung aus der Konjunktion von LOTH und CT (deren Unvereinbarkeit mit NTDP gezeigt werden soll) ist hier besonders hervorzuheben, da sie in Cummins' erstem Teilargument eine zentrale Rolle spielt:

Unter Annahme der minimalen Version von LOTH, die Cummins' zweitem Teilargument zufolge von der CT impliziert wird, folgt aus der CT für ein repräsentationales System, dass jede Eigenschaft, die von einem LOT-Grundterm in dem System repräsentiert wird, von dem System detektiert werden kann.

„[...] assuming LOT, CT requires that every property represented by a primitive in LOT be detectable.”<sup>12</sup> (Cummins 1997, S. 536).

Diese Folgerung aus CT und LOTH wird von Rupert in Rupert (2001) nicht in Frage gestellt.

## 2.2.2. Gelernte und native LOT Ausdrücke

### 2.3.1. Cummins' erster Argumentationsschritt

Cummins' erster Argumentationsschritt besteht darin zu begründen, dass es überhaupt eine nennenswerte Menge von LOT-Grundtermen gibt, deren MTs gelernt werden müssen.

Als paradigmatisches Beispiel für einen LOT-Grundterm, dessen MT gelernt werden muss, wird hier der LOT-Grundterm genannt, dessen Gehalt die Eigenschaft ist eine Katze zu sein. Dieser LOT-Grundterm wird im Folgenden, wie auch in Cummins (1997), mit |cat| bezeichnet.<sup>13</sup>

Cummins zufolge kann man die LOT-Grundterme in einem repräsentationalen System bezüglich der Herkunft ihrer MTs eindeutig in zwei wesentliche Klassen unterteilen.

„According to LOT, if it [die MT eines bestimmten LOT-Grundterms] is learned, it will be a set of sentences in LOT. If it is innate, it might, in some sense, be implicit in the architecture.” (Cummins 1997, S. 536, Fußnote 5).

Die eine Klasse von LOT-Grundtermen ist also dadurch ausgezeichnet, dass ihren jeweiligen MTs die Eigenschaft zukommt gelernt zu sein. Dies impliziert nach Cummins auch die Eigenschaft explizit als Menge von LOT-Sätzen repräsentiert zu sein.

Die andere zu ebengenannter komplementäre Klasse von LOT-Grundtermen ist dadurch ausgezeichnet, dass ihre jeweilige MT angeboren ist. Diese Eigenschaft ist Cummins zufolge für MTs von LOT-Grundtermen äquivalent mit der Eigenschaft nicht gelernt zu sein. Cummins vermutet, dass diese Eigenschaft für MTs von LOT-Grundtermen die Eigenschaften implizit repräsentiert zu sein und durch die Architektur des Systems realisiert zu sein impliziert<sup>14</sup>.

---

<sup>11</sup> Mit LOT ist hier LOTH gemeint. Siehe Fußnote 5.

<sup>12</sup> Die Abkürzung LOT bedeutet hier in der ersten Benutzung LOTH. Die zweite Nennung bezieht sich auf die „Language of Thought“ selbst (siehe Fußnote 5).

<sup>13</sup> Wird ein natürlichsprachliches Begriffswort in Betragsstriche gesetzt, bezieht sich dieser Ausdruck bei Cummins stets auf die Repräsentation, welche die Eigenschaft repräsentiert unter den Begriff zu fallen, welcher mit dem natürlichsprachlichen Wort bezeichnet wird. Die LOTH Annahme, dass eine mentale Repräsentation überhaupt Begriffe als Gehalt haben kann, setzt Cummins mit dieser Bezeichnung schon voraus. Nach seiner eigenen Theorie können Repräsentationen zwar Einzeldinge, Propositionen oder Wörter zum Inhalt haben, aber nicht Begriffe.

<sup>14</sup> Der Ausdruck, implizit repräsentiert zu sein, ist eigentlich ein Widerspruch. Repräsentiert zu sein bedeutet schließlich, der Gehalt einer echten, also expliziten Repräsentation zu sein. „Representing“ wird hier also metaphorisch gebraucht für Fälle, in denen einem Gegenstand Gehalt zugeschrieben werden kann, obwohl dieser Gegenstand bezüglich der angenommenen Theorie mentaler Repräsentation (CT, Bildtheorie, ...) nicht als Repräsentation bezeichnet werden kann. Da Cummins selber, insbesondere in Cummins (1986), „repräsentieren“ in dieser Weise benutzt (dort wird von „implicit representation“ gesprochen; S. 124) und auch Rupert von „implicit representation“ spricht, werde ich im Folgenden auch den Ausdruck „implizite Repräsentation“ gebrauchen.



Um zu begründen, dass |cat| ein LOT-Grundterm ist, dessen MT gelernt werden muss, argumentiert Cummins erstens dafür, dass |cat| überhaupt ein LOT-Grundterm ist und zweitens dafür, dass |cat| nicht in die Klasse von LOT-Grundtermen gehört, deren MTs angeboren und vermutlich implizit in der Architektur repräsentiert sind.

Um zu begründen, dass |cat| ein LOT-Grundterm ist, versucht Cummins es intuitiv unplausibel zu machen, dass |cat| ein komplexer LOT-Term ist.

„But if it [|cat|] is not a primitive symbol in LOT, then it must be a complex symbol in LOT, that is, a Tarskian combination of other symbols in LOT. This implies that |cat| can be defined in LOT in terms of symbols for properties whose detection does not require a learned theory. Good luck: this kind of reductionism has a dismal history; [...]” (Cummins 1997, S. 537).

Cummins Begründung dafür, dass die MT für |cat| nicht angeboren ist, und implizit ist, besteht in der Behauptung, dass grundlegendere Fähigkeiten als die Detektion der von |cat| repräsentierten Eigenschaft, wie die Fähigkeit zur Unterscheidung zwischen Kanten und Schatten, auch nicht angeboren sind.

„Much of the knowledge that mediates the detection of distal properties must be acquired: we are, perhaps, born with a tacit knowledge of Emmert’s law, but we are not born knowing how cats look, or with the ability to distinguish edges from shadows.” (Cummins 1997, S. 536 f.).

Aus diesen beiden Voraussetzungen lässt sich nach Cummins folgern, dass die MT für |cat| gelernt werden muss.

„We must, then, learn the theory that mediates cat recognition. Learning the theory will require formulating and confirming hypotheses such as these:

- (A) Cats have whiskers.
- (B) Cats have four legs.
- (C) Cats have fur.

According to LOT, these hypotheses are represented as sentences in LOT.” (Cummins 1997, S. 537)<sup>15</sup>.

|cat| steht stellvertretend für alle LOT-Terme, für welche es nicht plausibel erscheint, dass sie komplex sind, und deren Erzeugung durch Detektion gleich anspruchsvoll oder anspruchsvoller erscheint als die Unterscheidung von Kanten und Schatten.

Daraus, dass die MTs dieser LOT-Grundterme nicht angeboren sind lässt sich nach Cummins folgern, dass ihre MTs gelernt sind, und daraus wiederum, dass sie explizit repräsentiert sind.

Das Ergebnis des ersten Argumentationsschrittes ist also, dass unter Voraussetzung der all-gemeinsten CT Version |cat|s MT gelernt werden muss und explizit repräsentiert ist.

### **2.3.2. Kommentar**

Eine offenbare Schwäche von Cummins’ erstem Argumentationsschritt ist, dass die einzelnen Voraussetzungen der Folgerungsbeziehung nicht weiter begründet oder belegt werden.

Aus den in 2.2.2.1. zitierten Passagen lassen sich eine Disjunktionsbehauptung über Eigenschaften von LOT-Grundtermen sowie eine Implikationsbehauptung und eine Implikationsvermutung, welche eine klare Einordnung des Beispiels |cat| in eine der beiden disjunkten Mengen von LOT-Grundtermen suggeriert, rekonstruieren.

---

Es sei jedoch darauf hingewiesen, dass „Repräsentation“ im Zusammenhang mit „implizit“ metaphorisch zu verstehen ist.

<sup>15</sup> Mit der ersten Benutzung der Abkürzung „LOT“ meint Cummins die LOTH, mit der zweiten die „Language of Thought“ selbst (siehe Fußnote 5).

Des Weiteren wird eine Behauptung über den Status von |cat| als LOT-Grundterm gebraucht um im ersten Argumentationsschritt darauf zu schließen, dass die MT für |cat| gelernt werden muss.

Ich werde diese Behauptungen im Folgenden nacheinander besprechen.

### 2.3.2. (A) Implikationsvermutung

Cummins vermutet für LOT-Grundterme folgende Implikationsbeziehung:

Wenn der MT eines LOT-Grundterms die Eigenschaft zukommt angeboren zu sein, dann kommen der MT dieses LOT-Grundterms die Eigenschaften implizit zu sein und durch die Architektur des repräsentationalen Systems realisiert zu sein zu<sup>16</sup>.

In Rahmen seiner Kritik an Cummins' erstem Argumentationsschritt untersucht Rupert die eben genannte Implikationsvermutung sowie deren Bedeutung für die Einordnung des LOT-Grundterms |cat| in eine der disjunkten Klassen von LOT-Grundtermen. Ich diskutiere diese Vermutung deshalb ausführlich in 2.3.6.

Dasselbe gilt für Cummins' Behauptung, |cat|s MT wäre nicht angeboren.

### 2.3.2. (B) Implikationsbehauptung

Cummins behauptet in der Begründung seines ersten Argumentationsschrittes, dass für LOT-Grundterme folgende Implikationsbeziehung besteht:

Kommt der MT eines LOT-Grundterms die Eigenschaft zu gelernt zu sein, dann kommt der MT auch die Eigenschaft zu explizit repräsentiert zu sein.

Diese Implikationsbehauptung wird zwar von Rupert uneingeschränkt übernommen, scheint mir jedoch ebenso problematisch wie die eben genannte Implikationsvermutung.

Somit kritisiere ich diese Behauptung schon in diesem Kapitel.

Dazu werde ich folgende Thesen argumentativ belegen:

- Cummins' Behauptung ist nur unter Voraussetzung eines eingeschränkten Lernbegriffs nachvollziehbar.
- Diese Einschränkung des Lernbegriffs folgt nicht aus der minimalen Formulierung von CT und LOTH. Deshalb muss für die Behauptung ein allgemeinerer Lernbegriff angenommen werden.

---

<sup>16</sup> Cummins und Rupert bezeichnen die MT eines LOT-Grundterms mit dem bestimmten Artikel. Dies legt die Annahme nahe, dass jeder LOT-Grundterm genau eine MT hat. Da es in der LOT prinzipiell möglich ist den Inhalt eines Satzes auf verschiedene Weisen auszudrücken, ist diese Voraussetzung sogar für explizit repräsentierte MTs fragwürdig.

Eine, wenn auch etwas weniger starke, Voraussetzung muss allerdings gemacht werden:

Cummins und Rupert gehen beide stillschweigend davon aus, dass es prinzipiell möglich ist, die Menge der LOT-Grundterme durch Eigenschaften, welche ihren MTs zukommen, in disjunkte Klassen zu zerlegen, obgleich sie sich nicht einig sind, wie die Zerlegung genau aussieht.

Die Möglichkeit einer solchen Zerlegung erfordert für die Zuordnung eines LOT-Grundterms zu einer MT einige Einschränkungen:

Damit sich die Zerlegung der Menge der MTs durch Eigenschaften wie „implizit“ und „explizit“ auch auf die Menge der LOT-Grundterme überträgt, muss für jeden LOT-Grundterm gelten: Kommt eine klassifizierende Eigenschaft einer MT des LOT-Grundterms zu, so haben alle MTs des LOT-Grundterms diese Eigenschaft.

Diese Voraussetzung ist jedoch nicht durch die Minimalversionen von LOTH und CT zu begründen. Z.B. zeige ich in 2.3.2 (B), wie ein LOT-Grundterm eine explizite und eine implizite MT haben kann.

Somit ist die disjunkte Zerlegung der Menge von LOT-Grundtermen durch Eigenschaften ihrer MTs insgesamt in Frage zu stellen.

Für die weitere Darstellung von Cummins' Argument und Ruperts Kritik werde ich jedoch vorerst, so fragwürdig dies ist, davon ausgehen, dass es zwischen MTs und LOT-Grundtermen eine bijektive Zuordnung gibt. Dies wollen Cummins und Rupert vermutlich durch den Gebrauch des bestimmten Artikels andeuten.

- Unter Annahme eines allgemeinen Lernbegriffs gibt es mehrere mit LOTH und CT verträgliche Wege, auf denen eine implizite MT gelernt werden kann. Somit ist Cummins Implikationsbehauptung für einen allgemeineren Lernbegriff unzulässig.

### 2.3.2. (B) 1. Explizites Lernen

Die kurze aber sehr einschränkende Charakterisierung eines Lernvorgangs bei Cummins habe ich bereits zitiert:

„Learning the theory will require formulating and confirming hypotheses such as these:

- (D) Cats have whiskers.
- (E) Cats have four legs.
- (F) Cats have fur.

According to LOT, these hypotheses are represented as sentences in LOT.” (Cummins 1997, S. 537).

Offenbar meint Cummins, dass die LOTH für einen Lernprozess einer MT erfordert, dass explizite LOT-Sätze gebildet werden und dann durch Erfahrung überprüft werden.

Im Folgenden werde ich einen so zu beschreibenden Lernvorgang explizites Lernen nennen.

Warum Lernen von MTs, bei Geltung der CT und LOTH, als explizites Lernen beschrieben werden muss, dafür hat Cummins weder eine Begründung noch einen Verweis auf dahin gehende Forderungen von Vertretern der CT gegeben.

Da Cummins allerdings für andere Folgerungen aus CT und LOTH mehrfach Pylyshyn (1984) zitiert und Pylyshyn (ein Vertreter der LOTH; siehe 1.1.4. und 2.2.1.) bezüglich des Lernens von LOT-Termen eine ähnliche Position vertritt, liegt es nahe, dass Cummins bei der Formulierung seiner Beschreibung des Lernvorgangs durch Pylyshyns Argumentation beeinflusst ist.

Diese werde ich deshalb kurz skizzieren:

Lernen von MTs ist für Pylyshyn ein auf die Umwelt gerichteter Informationsverarbeitungsprozess. Ein solcher Prozess kann nur durch Operationen über Repräsentationen erklärt werden. Für Repräsentationen gilt jedoch die LOTH. Also kann ein Lernvorgang einer MT nur in der Überprüfung von LOT-Sätzen bestehen.

Wenn man Lernen mit explizitem Lernen gleichsetzt, dann ist Cummins’ Implikationsbehauptung durchaus gerechtfertigt. Wenn Lernen in der systematischen Formulierung und Überprüfung von expliziten LOT-Sätzen besteht, dann ist es naheliegend, die LOT-Sätze, deren „confirmation“ im letzten Lernschritt positiv verläuft, als Lernergebnis zu betrachten. Also wäre die gelernte MT explizit repräsentiert.

### 2.3.2. (B) 2. Lernen

Cummins beansprucht mit seinem Argument die Minimalversionen von CT und LOTH zu widerlegen. Wenn er für sein Argument die Identifikation von Lernen mit explizitem Lernen voraussetzt, muss diese Einschränkung des Lernbegriffs somit aus den Grundannahmen der CT und LOTH folgen.

Dies ist jedoch nicht der Fall. LOTH und CT sind durchaus mit einem weiter gefassten Lernbegriff verträglich.

Auch bei Pylyshyn ist die einschränkende Auffassung von Lernprozessen eher ein Folge dessen, dass er Lernprozesse für nicht anders als durch Repräsentationen beschreibbar hält, und keine direkte Folge aus der LOTH<sup>17</sup>.

In seinen anderen Kommentaren zur CT (z.B. Cummins 1996a, 2000, 2002b) benutzt Cummins meist einen weiter gefassten Lernbegriff. Lernen von Repräsentationen wird dort ganz allgemein als eine schrittweise Approximation an ein korrektes repräsentationales Verhalten anhand von Fehlersignalen beschrieben. So kritisiert Cummins die CT dort auch meist damit, dass sie keine für Lernen verwendbare Definition eines repräsentationalen Fehlers zulässt<sup>18</sup>.

Unter Annahme dieses allgemeineren Lernbegriffs ist Cummins' Implikationsbehauptung unzulässig, da es, wie ich im nächsten Abschnitt zeigen werde, mehrere mit LOTH und CT vereinbare Möglichkeiten gibt, implizit repräsentierte MTs zu lernen.

Da Cummins' Schluss im ersten Argumentationsschritt von der Richtigkeit der Implikationsbehauptung abhängig ist (siehe 2.3.2. (F)), kann Cummins' CT-Kritik zwar als Argument gegen eine CT/LOTH-Version, wie sie von Pylyshyn vertreten wird, gelten, nicht aber als schlüssiges Argument gegen die minimale Version der CT, wie sie in 2.2.3. definiert worden ist.

### 2.3.2 (B) 3 Lernen impliziter MTs

In diesem Abschnitt werde ich zwei Möglichkeiten skizzieren, wie MTs in einem LOTH/CT Schema gelernt und trotzdem implizit sein können.

Für die erste Variante wird der Begriff der Kodierung gebraucht, welchen Cummins in Cummins (1996b) und Cummins et al. (2001) im Zusammenhang mit der Systematizitätsdebatte diskutiert. Kodierungen ähneln Repräsentationen insofern, als dass sie ebenfalls einen intentionalen Gehalt haben können. Aber im Gegensatz zu Repräsentationen haben semantisch komplexe Kodierungen keine syntaktisch komplexe Struktur.

Es ist jedoch möglich, dass eine Kodierung Informationen über die Struktur ihres Gehaltes enthält ohne diese Struktur selbst zu haben. Ob es sich bei einer Kodierung um „struktural encoding“, also um Strukturinformation enthaltende Kodierung, oder um „pure encoding“ handelt, ist dadurch bestimmt, ob es für die Zuordnung von Kodierung zu strukturiertem Gehalt eine systematische Umkehrfunktion gibt<sup>19</sup>.

In Cummins et al. (2001) argumentiert Cummins gegen die LOTH, indem er zu zeigen versucht, dass durch ein Repräsentationsschema, für welches die LOTH gilt, nur sprachlich strukturierte Gehalte explizit repräsentiert werden können. Semantisch komplexe Gehalte, welche eine Struktur haben, die nicht zur LOT-Syntax isomorph ist<sup>20</sup>, können in einem

---

<sup>17</sup> Wie die Kategorisierung des Lernens als semantisch zu beschreibender Prozess, folgt nach Pylyshyn auch die LOTH selbst aus seinem, in 2.2.1. beschriebenen, Drei-Ebenen-Beschreibungsmodell repräsentationaler Systeme – und nicht umgekehrt.

<sup>18</sup> Dies kann hier natürlich nicht vorausgesetzt werden. Wenn Cummins hier annehmen würde, dass die CT Lernen wegen der Unmöglichkeit eines repräsentationalen Fehlers von vornherein nicht erklären kann, wäre der im zweiten Argumentationsschritt konstruierte Widerspruch trivial. Die CT wäre dann schon mit dem Ergebnis des ersten Argumentationsschrittes, dass bestimmte MTs gelernt werden müssen, widerlegt.

<sup>19</sup> Ein paradigmatisches Beispiel für eine strukturelle Kodierung ist bei Cummins das „Gödel numbering“: „[...] for encoding sentences, words are assigned natural numbers while their positions in the sentence are assigned prime numbers in ascending order. A number  $m$  in position  $n$  yields the number  $n^m$ . [...] The Gödel string of a sentence is determined by multiplying all these uniquely factorable numbers together.“ (Cummins et al. 2001, S. 183). Das Ergebnis der Gödel-Nummerierung – eine Zahl – besitzt offenbar nicht die syntaktische Struktur des kodierten Satzes, aber eine systematische Umkehrfunktion, sofern die den Wörtern zugeordneten Zahlen bekannt sind.

<sup>20</sup> Als Beispiel für eine zur LOT nicht isomorphe Struktur nennt Cummins unter anderen die räumlichen Relationen zwischen zu repräsentierenden Objekten: „[...] three-dimensional scenes are not isomorphic to Tarskian

LOT-Schema nicht komplex repräsentiert werden. Solche Gehalte können jedoch nach Cummins strukturell kodiert werden. Der Gehalt eines LOT-Grundterms könnte dann, relativ zu einem anderen (zur LOT nicht isomorphen) syntaktischen Schema, komplex sein. Und der LOT-Term könnte, im Falle einer strukturellen Kodierung, obwohl er ein Grundterm ist, Informationen über die Struktur des Gehaltes enthalten ohne diese Struktur selber zu haben.

Mit diesen Begriffen kann man einen impliziten Lernvorgang bei Geltung der LOTH und CT folgendermaßen beschreiben:

Sei  $|a|$  ein LOT-Grundterm, dessen MT aus Sätzen in einem syntaktischen Schema besteht, welches zur LOT nicht isomorph ist<sup>21</sup>.  $|a|$ 's MT ist also implizit.

Diese MT kann nun dadurch gelernt werden, dass das repräsentationale System Kodierungen von Sätzen des (zur LOT nicht isomorphen) syntaktischen Systems formuliert und diese dann überprüft. Die Kodierung, für welche die Überprüfung im letzten Lernschritt erfolgreich verläuft, ist  $|a|$ 's MT.

Diese Beschreibung eines Lernprozesses wirft einige Fragen auf, die ich im Folgenden kurz behandeln werde:

- Wie können Kodierungen formuliert werden?
- Wie können Kodierungen überprüft werden?
- Ist ein solcher Lernvorgang im Rahmen der LOTH überhaupt sinnvoll vertretbar?

Zur Formulierung von Kodierungen:

Als Beispiel eines Kodierungsalgorithmus wird in Cummins et al. (2001) ein in Smolensky et al. (1992) beschriebener Parser, welcher Sätze mit einer sprachähnlichen Syntax durch Tensorprodukte kodiert, angeführt. Dieser Parser kann Sätze nicht nur gemäß ihrer Syntax klassifizieren, sondern auch wohlgeformte Sätze generieren. Das spricht dafür, dass es möglich ist in einem repräsentationalen System Kodierungen von Sätzen zu formulieren, deren Syntax nicht der Syntax des repräsentationalen Systems entspricht.

Zur Überprüfung von Kodierungen:

Lernen im oben beschriebenen Sinne ist Lernen mit Fehlersignal, welches entsprechend einem repräsentationalen Fehler gebildet wird. Die Überprüfung („confirming“) besteht also in dem Vergleich von gebildeter Repräsentation mit der korrekten Repräsentation.

Das „tokening“ einer Repräsentation mittels einer kodierten MT kann somit als Überprüfung der MT angesehen werden.

Cummins hält den Begriff „repräsentationaler Fehler“ für LOT-Repräsentationen zwar aus allgemeineren Gründen (in z.B. Cummins 1996a, 2000) für unmöglich, aber dieser Einwand kann hier nicht gelten, da er auf explizites Lernen genau so anzuwenden wäre (siehe auch Fußnote 18).

Zur Vereinbarkeit mit der LOTH:

In Cummins (1996b) und Cummins et al. (2001) führt Cummins die Möglichkeit der Kodierung eigentlich gegen die LOTH ins Feld:

Wenn es in unterschiedlichen zu repräsentierenden Bereichen („domains“) unterschiedliche nicht zueinander isomorphe Systematizitäten gibt, und wenn in einer LOT angeordnete Repräsentationen Gehalte, welche einer zur LOT-Syntax nicht isomorphen Syntax unterliegen, nur kodieren, aber nicht repräsentieren können, dann kann das Systematizitätsargument nicht

---

structures, [...]. No scheme can structurally represent both sentence structure and spatial structure, since these are not isomorphic to each other.” (Cummins 1996b, S. 604).

<sup>21</sup> Naheliegender wäre das für die Repräsentation einer Eigenschaft, welche sich als räumliche Relation ausdrücken lässt (siehe Fußnote 20). Ob man  $|a|$ 's MT zusätzlich auch LOT-explizit ausdrücken kann, ist für dieses Argument nicht relevant (siehe Fußnote 16).

mehr für die LOTH sprechen (siehe Kapitel 1.2.3.). Das Systematizitätsargument würde dann von einer zur LOT nicht isomorphen Systematizität in einigen zu kognizierenden Bereichen auf ein von der LOT unterschiedenes repräsentationales Schema schließen, würde also gegen die LOTH sprechen.

In der Tat ist es wenig plausibel, dass jemand die LOTH vertritt und gleichzeitig annimmt, dass ein großer Teil mentaler Gehalte nicht repräsentiert, sondern kodiert wird.

Theoretisch hält Cummins so ein Modell jedoch ausdrücklich für möglich. Deshalb ist es hier auch angebracht ein solches Modell zu argumentativen Zwecken zu gebrauchen. Cummins möchte nämlich mit dem hier zu untersuchenden Argument nicht eine bestimmte aktuell vertretene Version der CT/LOTH widerlegen, sondern zeigen, dass die CT aus prinzipiellen Gründen falsch ist.

Dass es eine theoretische mit LOTH und CT vereinbare Möglichkeit gibt implizite MTs zu lernen, zeigt, dass die Implikationsbehauptung nicht aus den Minimaldefinitionen von LOTH und CT folgt, dass die Implikationsbehauptung für die CT also nicht allgemein gilt.

Eine andere Möglichkeit implizit repräsentierte MTs zu lernen, ist, einen Lernvorgang durch systematische Veränderung der funktionalen Architektur zu beschreiben. Cummins vermutet in seiner Implikationsvermutung, dass implizite MTs bei Geltung der CT und LOTH grundsätzlich in der funktionalen Architektur realisiert sind.

Es ist jedoch denkbar, dass Veränderungen der funktionalen Architektur durch eine fehlerhafte Bildung<sup>22</sup> eines LOT-Grundterms verursacht werden können und somit gemäß dem oben genannten allgemeinen Lernbegriff, als Lernvorgänge beschrieben werden. Solche Prozesse könnten, wie auch Clapin vorschlägt, als Informationsverarbeitungsprozesse beschrieben werden, ohne dass sie explizite Repräsentationen enthalten<sup>23</sup>: „[...] changes to functional architecture ought be understood, in certain circumstances, as a kind of cognition.“ (Clapin 2002, S. 306)

Dies wäre damit vereinbar, dass für das Schema der expliziten Repräsentationen die LOTH gilt.

Es ist hier also festzuhalten, dass Cummins' Implikationsbehauptung nicht aus der allgemeinen Version von CT und LOTH folgt.

### 2.3.2. (C) Disjunktionsbehauptung

Die Disjunktionsbehauptung wird von Cummins in Cummins (1997) nicht eindeutig formuliert.

Dies liegt wohl daran, dass Cummins seinen ersten Argumentationsschritt äußerst knapp, teilweise in Fußnoten, formuliert und diesem in seinen Repliken auf potentielle Einwände auch keine weitere Beachtung mehr schenkt.

Die Sätze „According to LOT, if it [die MT eines bestimmten LOT-Grundterms] is learned, it will be a set of sentences in LOT. If it is innate, it might, in some sense, be implicit in the architecture.“ (Cummins 1997, S. 536, Fußnote 5) könnten ohne Rücksicht auf den Kontext auch lediglich als Aneinanderreihung einer Implikationsbehauptung und einer Implikationsvermutung verstanden werden.

---

<sup>22</sup> Wieder wird hier vorausgesetzt, dass Fehlersignale bei Geltung der CT möglich sind (siehe Fußnote 18).

<sup>23</sup> Nach Pylyshyns Modell wäre so etwas natürlich nicht möglich. Wenn Änderungen der funktionalen Architektur als Kognition beschrieben werden, handelt es sich per Definition nicht mehr um funktionale Architektur. „Some of these mechanisms may be quite rigid, whereas others may be alterable in various ways by the environment. [...] Of course, the mechanisms themselves are not sensitive to the *content* of incoming information, since, by hypothesis, semantic content is precisely what is encoded in terms of the symbolic codes[...].“ (Pylyshyn 1984, S. 30 f.). Aus der LOTH folgt dies jedoch nicht (siehe Abschnitt 2.3.2. (B)).

Der Kontext legt jedoch nahe, dass Cummins hier – neben den Implikationsbeziehungen – ausdrücken möchte, dass der MT eines jeden LOT-Grundterms in einem repräsentationalen System entweder die Eigenschaft zukommt gelernt zu sein oder die Eigenschaft zukommt angeboren zu sein.

Anderenfalls wäre der Schluss in Cummins' erstem Argumentationsschritt offensichtlich ungültig. Er kann nur dann von der Aussage, dass  $|cat|_s$  MT nicht angeboren ist, direkt auf die Aussage schließen, dass  $|cat|_s$  MT gelernt ist, wenn die ebengenannte Disjunktionsbehauptung gilt.

So rekonstruiert Rupert die Disjunktionsbehauptung dementsprechend und baut darauf einen wesentlichen Teil seiner Kritik auf (siehe 2.3.3. und 2.3.5.).

### 2.3.2. (D) $|cat|$ ist ein LOT-Grundterm

Die Möglichkeit, dass  $|cat|$  kein LOT-Grundterm ist, wird bei Cummins von vornherein ausgeschlossen. Dabei verweist er auf die Erfolglosigkeit von Ansätzen zur definitorischen Reduktion solcher Terme.

„But if it [ $|cat|$ ] is not a primitive symbol in LOT, then it must be a complex symbol in LOT, that is, a Tarskian combination of other symbols in LOT. This implies that  $|cat|$  can be defined in LOT in terms of symbols for properties whose detection does not require a learned theory. Good luck: this kind of reductionism has a dismal history; [...]“ (Cummins 1997, S. 537).

Weitere Hinweise darauf, worum es sich bei dieser „dismal history“ handelt, sind in Cummins' Text nicht zu finden.

Als mögliche Bezugspunkte für Cummins' Intention kommen mehrere, nicht mehr weiter verfolgte Projekte definitorischer Reduktion in Frage<sup>24</sup>:

Z.B. der Versuch der logischen Empiristen theoretische Ausdrücke und Objektbezeichnungen auf Wahrnehmungs-Sprache zu reduzieren<sup>25</sup>. Oder das sogenannte „semantic-markers“ Projekt, welches auch in Fodor (1998) kritisiert wird.

Diese Beispiele reichen aus um Cummins in der Einschätzung Recht zu geben, dass der Versuch, Terme wie  $|cat|$  definitorisch auf angeborene Terme zu reduzieren, wenig erfolgversprechend ist, und  $|cat|$  hier deshalb als Grundterm aufzufassen ist.

### 2.3.2. (E) Der Schluss

Für den ersten Argumentationsschritt setzt Cummins voraus, dass folgende vier Behauptungen bei Annahme der CT und LOTH gelten:

- (1)  $|cat|$  ist ein LOT-Grundterm.
- (2)  $|cat|_s$  MT ist nicht angeboren.
- (3) Für MTs von LOT-Grundtermen gilt: Eine MT ist entweder angeboren oder sie ist gelernt.
- (4) Für MTs von LOT Grundtermen gilt: Wenn eine MT gelernt ist, ist sie explizit repräsentiert.

Cummins schließt zuerst von (1), (2) und (3) darauf, dass  $|cat|_s$  MT gelernt ist. Rupert argumentiert in seiner Kritik an Cummins' erstem Argumentationsschritt dafür, dass entweder (2) oder (3) falsch ist, wobei er (1) und (4) akzeptiert.

Wenn sich Ruperts Kritik als berechtigt herausstellen sollte, kann Cummins nicht mehr darauf schließen, dass  $|cat|_s$  MT gelernt ist.

Aus der Aussage, dass  $|cat|_s$  MT gelernt ist, schließt Cummins mit (4) darauf, dass  $|cat|_s$  MT gelernt und explizit ist.

---

<sup>24</sup> Cummins hat diese Interpretation der „dimal history“ mir gegenüber mündlich bestätigt.

<sup>25</sup> In diesen Ansätzen wurden zwar überhaupt keine mentalen Repräsentationen angenommen. Trotzdem wäre der Versuch, LOT-Terme definitorisch zu reduzieren mit den selben Problemen konfrontiert.

Auch wenn man die Aussage, dass |cat|s MT gelernt werden muss, trotz Ruperts Kritik akzeptiert, ist Cummins' Schluss in Frage zu stellen, da ich die Aussage (4) in Abschnitt 2.3.2. (B) widerlegt habe.

Man kann also insgesamt nicht davon ausgehen, dass aus der Minimaldefinition von CT und LOTH folgt, dass |cat|s MT explizit und gelernt ist.

### **2.3.3. Cummins' erster Argumentationsschritt bei Rupert**

Um die Angriffspunkte seiner Kritik herauszustellen, formuliert Rupert in Rupert (2001) eine Rekonstruktion von Cummins' erstem Argumentationsschritt. Ich werde diese Rekonstruktion hier kurz darstellen. Rupert beschreibt die Disjunktionsbehauptung, die Behauptung, dass |cat|s MT nicht angeboren ist, und den daraus resultierenden Schluss so:

„(P1) CT can only account for the fixation of the content of the primitive LOT term 'cat' by appealing either (a) to what the human subject know at birth about cats or (b) to the human subject's acquisition of an explicit theory of cats.

(P2) What the human subject knows at birth does not fix the content of her primitive LOT term 'cat'.

Therefore, an explicit theory of cats must be acquired, if CT provides the correct account of how the human subject acquires the primitive LOT term 'cat'.”<sup>26</sup> (Rupert 2001, S. 504).

Die Implikationsbehauptung reformuliert Rupert so: „Let us use *learned terms* to refer to these LOT terms for which, assuming CT and NTDP, it is a necessary condition of their acquiring content that the human subject formulate an explicit theory of that to which the terms apply; [...].”(Rupert 2001, S. 502).

Die Implikationsvermutung (siehe 2.2.2.2. (A)) wird von Rupert nicht reformuliert. In deren Kritik bezieht er sich direkt auf Cummins Formulierung und zitiert diese auch in den betreffenden Passagen.

Cummins Behauptung, dass |cat| ein LOT-Grundterm ist, folgt Rupert uneingeschränkt.

„Following Cummins, I set aside the possibility that 'cat' is compound.”(Rupert 2001, S. 504, Fußnote 5).

### **2.3.4 Kommentar**

In Ruperts Rekonstruktion erscheinen einige von Cummins' Aussagen in inhaltlich leicht veränderter Form. Zwei Veränderungen werde ich hier kurz kommentieren.

Die Implikationsbehauptung von Cummins wird bei Rupert zu einer Äquivalenzaussage verschärft.

Diese Interpretation ist insofern gerechtfertigt, als dass Cummins seine Implikationsvermutung – angeboren zu sein impliziert, implizit repräsentiert zu sein – wohl kaum aufgestellt hätte, wenn er angenommen hätte, es gäbe MTs, die sowohl angeboren – also nach Disjunktionsbehauptung nicht gelernt – als auch explizit repräsentiert sind.

Die Behauptung von Cummins, dass die MT für |cat| nicht angeboren ist, erscheint bei Rupert als Satz (P2) .

Cummins' Disjunktionsbehauptung taucht in Ruperts Rekonstruktion als Satz (P1) auf.

---

<sup>26</sup> Rupert benutzt zur Kennzeichnung von LOT-Repräsentationen anstatt von Betragsstrichen einfache Anführungszeichen. Mit „'cat'“ meint Rupert also „|cat|“.



Einerseits wird in dieser Formulierung deutlicher als bei Cummins ausgedrückt, dass die LOT-Grundterme Cummins zufolge durch die entsprechenden Eigenschaften ihrer MTs bei Annahme der CT in zwei disjunkte Klassen zu unterteilen sind.

Diese Präzisierung trägt zur Verdeutlichung von Cummins' Argumentationsweise im ersten Argumentationsschritt bei und bietet Rupert einen eindeutig formulierten Bezugspunkt für seine Kritik. Dass diese Präzisierung inhaltlich gerechtfertigt ist, habe ich in 2.3.2. (C) dargelegt.

Andererseits „überspringt“ Rupert in seiner Rekonstruktion Cummins' Implikationsbehauptung und setzt als zweites Argument der Disjunktionsbehauptung nicht die Eigenschaft, gelernt zu sein, ein, sondern gleich die Eigenschaft, explizit repräsentiert zu sein, welche nach Cummins schließlich erst durch die Eigenschaft, gelernt zu sein, impliziert wird.

Für Ruperts Kritik an Cummins' Argument ist diese Veränderung jedoch nicht relevant, da Rupert die Implikationsbehauptung ohnehin nicht kritisiert oder sonst irgendwo weiter erwähnt.

### **2.3.5. Ruperts Kritik an Cummins' erstem Argumentationsschritt**

Ruperts Kritik an Cummins' erstem Argumentationsschritt besteht hauptsächlich in der Behauptung, dass der Begriff „angeboren“<sup>27</sup> im Gebrauch in Satz (P1) – bezogen auf die Menge der MTs von LOT-Grundtermen in einem repräsentationalen System – nicht extensionsgleich ist mit dem Begriff „angeboren“, wie er in Satz (P2) gebraucht wird, und in einer ausführlichen Begründung dieser Behauptung.

Da der Schluss in Cummins' erstem Argumentationsschritt erfordert, dass der Begriff „angeboren“ in (P1) und (P2) derselbe ist, wäre Cummins erster Argumentationsschritt bei Geltung dieser Behauptung ungültig.

Rupert zufolge würde die Disjunktionsbehauptung (P1) um wahr zu sein eine schwache Interpretation von „angeboren“ erfordern.

Unter einer solchen Interpretation von „angeboren“ wäre dieser Begriff in der Menge der MTs von LOT-Grundtermen eines repräsentationalen Systems<sup>28</sup> negativ definiert, als genau die MTs einschließend, denen nicht die Eigenschaften gelernt zu sein und explizit repräsen-

---

<sup>27</sup> Rupert trifft im Verlauf seiner Argumentation eine Unterscheidung zwischen einer starken und einer schwachen Bedeutung von „angeboren“. Die Wörter „innate“ und „native“ werden von Rupert manchmal für die starke Version von „angeboren“ benutzt, öfter aber als bezüglich dieser Unterscheidung neutral. Ist eine MT im starken Sinne angeboren, wird diese Theorie als „knowledge at birth“ bezeichnet. Manchmal wird „knowledge at birth“ allerdings, wie auch in diesem Satz, neutral gebraucht.

Ich werde der Eindeutigkeit halber jeden Gebrauch von „angeboren“, welcher bezüglich der obigen Unterscheidung nicht neutral ist, durch einen entsprechenden Zusatz kennzeichnen.

„Knowledge at birth“ werde ich im Folgenden mit „im starken Sinne angeborene MT“ übersetzen. Dieser Satz ist davon ausgenommen, da „knowledge at birth“ hier ausnahmsweise im neutralen Sinne benutzt wird. Die Übersetzung mit „angeboren“ erscheint hier auf den ersten Blick unglücklich, da es korrekter „angeborene MT“ heißen müsste.

Damit greife ich jedoch Ruperts Ausdrucksweise vor, welche später nach und nach von „knowledge at birth“ zu „different interpretations of innate“ übergeht.

<sup>28</sup> Für alle Aussagen über die Eigenschaften gelernt zu sein, ungelernet zu sein, explizit zu sein, implizit zu sein, in irgendeinem Sinne angeboren zu sein und Kombinationen und Modifikationen solcher Eigenschaften sei für alle folgenden Unterkapitel von II. vorausgesetzt:

Die Aussagen beziehen sich, falls kein anderer Bezug kenntlich gemacht ist, auf Eigenschaften von MTs von LOT-Grundtermen in einem beliebig fest zu wählenden repräsentationalen System unter Geltung der CT und LOTH in ihren in 2.2. genannten Definitionen.

tiert zu sein zukommen. Die, der so definierten Klasse von MTs entsprechende, Klasse von LOT-Grundtermen nennt Rupert ungelernete Terme.

„Let us use *learned terms* to refer to those LOT terms for which, assuming CT and NTDP, it is a necessary condition of their acquiring content that the human subject formulate an explicit theory of that to which the terms apply; call all other LOT terms *unlearned*.“ (Rupert 2001, S. 502). Da Cummins' Implikationsbehauptung bei Rupert als Äquivalenzbehauptung rekonstruiert und nicht in Frage gestellt wird, bezeichnet Rupert ungelernete Terme auch als Terme mit impliziter MT.

Für jede stärkere Definition von „angeboren“ wäre die Disjunktionsbehauptung natürlich falsch, da es dann ja LOT-Grundterme gäbe, deren MTs weder gelernt und explizit repräsentiert noch im stärkeren Sinne angeboren wären.

Da Rupert es jedoch ausdrücklich für möglich hält, dass |cat| in die Klasse der ungelernenen Terme fällt, muss „angeboren“ für (P2) in einem stärkeren Sinne interpretiert werden.

Falls „angeboren“ in (P2) in der schwachen Version als nicht gelernt und nicht explizit repräsentiert verstanden würde, wäre (P2) falsch, da Rupert zufolge |cat| in die Klasse der ungelernenen Terme gehören könnte, was in (P2) eben gerade bestritten wird.

Die starke Interpretation von „angeboren“ ist die, dass eine MT, die als „angeboren“ bezeichnet werden kann, tatsächlich bei der Geburt des das repräsentationale System enthaltenden Organismus schon funktionsfähig vorhanden ist.

Für diese Definition wäre die Behauptung (P2) nach Rupert richtig, da er die starke Definition von „angeboren“ bezüglich der MT von |cat| nicht für zutreffend hält.

Seine kritische Behauptung, dass es LOT-Grundterme gibt, auf welche die schwache, |cat|s MT möglicherweise einschließende, Definition von „angeboren“ zutrifft, die starke Definition von „angeboren“ jedoch nicht zutrifft, stützt Rupert mit mehreren Argumenten.

Zuerst versucht Rupert zu zeigen, dass Cummins' Implikationsvermutung – für MTs impliziert angeboren zu sein implizit in der Architektur repräsentiert zu sein – eine Definition von „angeboren“, also „knowledge at birth“, zulässt, die über die starke Definition hinausgeht. Um dies zu zeigen, argumentiert er dafür, dass es MTs gibt, die zwar implizit in der Architektur repräsentiert sind, aber sich nach der Geburt verändern können ohne in die Klasse der MTs von gelernten Termen zu fallen; also MTs, die im schwachen Sinne angeboren und in der Architektur implizit sind, jedoch nicht im starken Sinne angeboren sind.

Des Weiteren argumentiert Rupert dafür, dass es MTs gibt, die im schwachen Sinne angeboren sind, aber nicht in der Architektur repräsentiert sind.

Ergänzend dazu untersucht Rupert einige in den Kognitionswissenschaften übliche Definitionen von „angeboren“ daraufhin, ob sie der schwachen Definition von „angeboren“ gerecht werden.

Schließlich diskutiert er einige entwicklungspsychologische Paradigmen im Hinblick auf die Möglichkeit der Entstehung von MTs ohne explizite Lernvorgänge.

Aus diesen Untersuchungen schließt Rupert, dass, je nach Definition von „angeboren“, entweder Cummins' Disjunktionsbehauptung falsch ist (starke Definition) oder Cummins' Aussage, dass |cat|s MT angeboren ist, falsch ist (schwache Definition).

Somit wäre Cummins' Schluss im ersten Argumentationsschritt für |cat| ungültig.

Ich werde Ruperts Argumente nacheinander referieren und in 2.3.6. kommentieren.

### 2.3.5. (A) Plastizität funktionaler Architektur

In seiner ersten Analyse ungelernter MTs diskutiert Rupert, was die starke Definition von „angeboren“ für MTs bedeutet, welche in der Architektur enthalten sind, und stellt in Frage, dass alle MTs, die in der Architektur enthalten sind, in diesem Sinne „angeboren“ sind.

Die starke Definition von „angeboren“ muss für MTs, die in der Architektur realisiert sind, bedeuten, dass sich diese Architektur, zumindest sofern sie eine MT realisiert, nach der Geburt nicht verändert. Sonst würde sich auch die entsprechende MT ändern und die Eigenschaft nicht mehr korrekt detektieren.

Cummins ist offenbar der Auffassung, dass aus CT und LOTH folgt, dass diese Eigenschaft angeboren zu sein – mit „angeboren“ im starken Sinne – auf alle MTs zutrifft, die implizit in der Architektur enthalten und nicht gelernt sind. Dies würde bedeuten, dass sich die funktionale Architektur in einem repräsentationalen System nicht verändern kann.

Dafür, dass aus CT und LOTH diese Aussage folgt, gibt Cummins selbst keine Gründe an, verweist jedoch auf Pylyshyn (1984).

Wie in 2.2.1. bereits dargestellt wurde, hält Pylyshyn zur Beschreibung kognitiver Systeme drei Ebenen für notwendig:

Die physikalische, die symbolische und die semantische Ebene. Die semantische Ebene entspricht der Ebene expliziter Repräsentationen. Auf dieser Ebene werden kognitive Prozesse beschrieben, für deren Erklärung Intentionalität erforderlich ist.

Die symbolische Ebene entspricht der funktionalen Architektur eines repräsentationalen Systems. Auf dieser Ebene werden Mechanismen beschrieben, welche die Symbole ohne Bezug auf ihren repräsentationalen Gehalt verarbeiten.

Auf der physikalischen Ebene wird beschrieben, wie ebengenannte Mechanismen physikalisch realisiert sind.

Cummins scheint nun anzunehmen, dass Pylyshyn funktionale Architektur als unveränderbar auffasst. Nahegelegt wird dies durch die zitierte Passage: „[...] the architecture must form a cognitive ‘fixed point’ so that differences in cognitive phenomena can be explained by appeal to arrangements [...] among the fixed set of operations and to the basic resources provided by the architecture [...]” (Pylyshyn 1984, S.114).

Dies ist jedoch nach Rupert nicht so zu interpretieren, dass die funktionale Architektur sich nach Pylyshyn in einem repräsentationalen System nicht verändern kann, sondern als Forderung für die Erklärung kognitiver Prozesse auf der semantischen Ebene. Diese Prozesse dürfen nach Pylyshyn keine Veränderungen der funktionalen Architektur beinhalten.

Cummins’ Identifikation von MTs, die angeboren im starken Sinne, also unveränderbar, sind, mit MTs, die in der funktionalen Architektur realisiert sind (bei Geltung von CT und LOTH), beruht somit nach Rupert auf einer Fehlinterpretation von Pylyshyns Beschreibungsebenen.

„The sense in which Pylyshyn wishes to hold fixed the functional architecture will not, however, do the duty Cummins expects of it. Pylyshyn holds functional architecture fixed for explanatory purposes. [...] The preceding picture of functional architecture and its role as an explanatory fixed point does not require that each aspect of the functional architecture be either innate or unchanging.”(Rupert 2001, S. 506).

Schließlich nennt Rupert einige Möglichkeiten der Veränderung von funktionaler Architektur bei biologischen repräsentationalen Systemen, die auch bei Pylyshyn (1984) zu finden sind.

„Such modulations may be due to biochemical influences, maturation, the triggering effect of environmental releasers, dendritic arborisation, atrophy of neural function from disuse, and so on.“ (Pylyshyn 1984, S. 259).

### 2.3.5. (B) Funktionale Architektur und implizite MTs

Obwohl Rupert der Implikationsvermutung aus Cummins (1997), dass MTs, die implizit sind, auch in der funktionalen Architektur enthalten sind, ein eigenes Kapitel widmet (Rupert 2001, S. 509-510), gibt er keinerlei positive Argumente gegen diese Vermutung an, sondern begnügt sich mit einer variantenreichen Wiederholung der Bemerkung, dass Cummins seine Implikationsvermutung nicht begründet hat und die Klasse in der Architektur impliziter MTs nicht näher beschreibt. Z.B.: „[...] if he has in mind a substantive constraint on what counts as implicit theory built into the functional architecture, he should explicitly state the criterion and argue for it.“ (Rupert 2001, S. 509).

### 2.3.5. (C) „angeboren“-Definitionen bei Rupert

Als weitere Stütze seiner Behauptung, dass der Begriff „angeboren“ in (P1) und (P2) unterschiedlich große Klassen von MTs bezeichnen muss, zeigt Rupert, dass es in den Kognitionswissenschaften durchaus übliche Definitionen des Begriffs „angeboren“ gibt, die wesentlich schwächer sind, als die Definition, die Cummins für (P2) voraussetzen muss.<sup>29</sup>

Dies geschieht, indem Rupert einige Definitionen von „angeboren“ vorstellt und überprüft, ob diese über die starke Version von „angeboren“ hinausgehen, die für (P2) erforderlich wäre: „To what extent does any one of these conceptions of innateness imply that, for a subject to possess an innate concept, she must possess knowledge at birth about that to which the concept applies?“ (Rupert 2001, S. 511).

Dass der Begriff „angeboren“ in den Kognitionswissenschaften häufig in schwächerer Bedeutung benutzt wird, als im Sinne von „bei der Geburt vorhanden“, wertet Rupert als Beleg für die Möglichkeit von MTs, die in einem solchen schwächeren Sinne angeboren, im stärkeren Sinne jedoch nicht angeboren sind.

Da die Frage, inwiefern die besprochenen Definitionen von „angeboren“ tatsächlich alle ungelernen MTs einschließen könnten, für die Beurteilung von Ruperts Kritik, wie ich in 2.3.6. (C) darlegen werde, keine wesentliche Bedeutung hat, werde ich mich in der nun folgenden Darstellung der Definitionen sowie im Kommentar (s.u.) nur auf die Punkte der jeweiligen Definition beziehen, die Rupert explizit zitiert.

### 2.3.5. (D) Die Definitionen

Die erste Definition von „angeboren“ ist von Rupert aus Elman et al. (1996) entnommen worden und besagt, dass ein Verhalten oder ein Merkmal genau dann als „angeboren“ bezeichnet werden sollte, wenn dessen Realisierung unter normalen Entwicklungsbedingungen sehr wahrscheinlich ist.

Die Position wird in Rupert (2001) so zitiert: „When we say that a form or a behaviour is innate we really mean that given normal developmental experiences, it is a highly probable outcome“ (Elman et al 1996, S. 319).

---

<sup>29</sup> Rupert nimmt hier offenbar an, dass auch schwache Definitionen von „angeboren“ gelernte MTs ausschließen (siehe Abschnitt 2.3.6. (C)).

Rupert bemerkt, dass diese Definition von „angeboren“ zur Formulierung nativistischer Behauptungen sehr nützlich ist und dass diese Definition nicht erfordert, dass die angeborene Eigenschaft bereits bei der Geburt vorhanden ist. Die Definition ist also schwächer als die, welche nach Rupert von (P2) erfordert wird.

In derselben Quelle schlagen Elman et al. eine weitere Definition von „angeboren“ vor, die von Rupert ebenfalls besprochen wird. Nach dieser Definition sollte „angeboren“ folgendermaßen benutzt werden: „[...] to refer to putative aspects of brain structure, cognition or behaviour that are the product of interactions internal to the organism.“ (Elman et al. 1996, S. 23).

Auch diese Definition von „angeboren“ ist nach Rupert schwächer als die, welche durch (P2) erfordert wird.

„Here again, innateness does not require (P2)-knowledge. An innate concept must emerge from organismically internal relations, but this implies no particular form of knowledge present in the newborn supporting this emergence.“ (Rupert 2001, S. 512).

Die dritte Definition stammt von R. Samuels. Aus Samuels (1998) zitiert Rupert: „A characteristic C is genetically determined for an organism O just in case organisms that have the same combination of alleles as O develop C in all standard environments“ (Samuels 1998, S. 564).

Auch diese Definition beurteilt Rupert, wie die beiden vorhergehenden, als die starke Definition von „angeboren“ erweiternd.

„[...] it [die Definition] requires nothing in the way of knowledge, and certainly not (P2)-knowledge, in the newborn in order that a trait, behaviour, or concept be innate.“ (Rupert 2001, S. 514)

Die letzte von Rupert erwähnte Definition von „angeboren“ ist die von J. Fodor. Diese Definition beschreibt Rupert so: „According to this view, a concept is innate if and only if it is primitive, where being primitive now means *not being acquired by hypothesis testing and confirmation*.“ (Rupert 2001, S. 514). Rupert identifiziert diese Definition mit dem Gebrauch von „angeboren“ bei Cummins in (P1).

„Cummins`s discussion of innateness contrasts those LOT terms relative to which explicitly represented theory mediates content fixation and LOT terms relative to which it does not. By identifying innate terms with those in the latter class, Cummins seems to be following Jerry Fodor; [...]“ (Rupert 2001, S. 514).

### 2.3.5. (E) Entwicklungspsychologische Paradigmen

Als weiteres Argument für die Existenz von nicht gelernten und nicht im starken Sinne angeborenen MTs führt Rupert die Zusammenfassung einer gegenwärtigen Debatte in der Theorie neuronaler Entwicklung an.

In dieser Debatte geht es darum, ob Interaktion eines Organismus mit seiner Umwelt das Wachstum von Neuronen beeinflussen kann, oder lediglich die funktionale Organisation von schon existierenden Neuronen verändert.

„The interesting debate here is not over whether experience affects neural development – the experimental evidence seems to have settled that issue – but over the precise role of experience in shaping the structure of the brain.“ (Rupert 2001, S. 518).

Bezüglich dieser Frage schildert Rupert zwei entgegengesetzte Ansichten. Auf der einen Seite steht der „selektionistische“ Ansatz, welcher von Rupert hauptsächlich Edelman (1987) und Changeaux (1997) zugeschrieben wird.

„According to selectionist views, genetically programmed growth provides a wealth of raw material that experience, among other factors, then winnows to create functional neural structure.” (Rupert 2001, S. 518).

Dem gegenüber stellt Rupert den „konstruktivistischen” Ansatz: „They [konstruktivistische Ansätze] assign a further role to experience, however: it guides the generation of axonal and dendritic arbors and can selectively increase the number of synapses in an given area of the brain of a given neural circuit; [...]“ (Rupert 2001, S. 519).

Entscheidend ist hier für Rupert, dass in dieser Debatte vorausgesetzt wird, dass Interaktion mit der Umwelt überhaupt für die Veränderung der neuronalen Architektur verantwortlich ist. Dies zeigt nämlich gemäß Rupert, dass die neuronalen Träger von Repräsentationen zum größten Teil nicht im starken Sinne „angeboren“ sind.

„If selectionists *or* constructivists are correct, the vast majority, perhaps all, of an adult`s nonsemantically individuated LOT terms appear in no substantial form at birth.” (Rupert 2001, S. 520).

### 2.3.5. (F) Cummins` Schluss auf die Notwendigkeit des Lernens von |cat|s MT

In diesem Kapitel soll dargestellt werden, wie Rupert die Wirkung seiner Argumentation auf Cummins` ersten Argumentationsschritt beurteilt.

Cummins schließt von der Aussage, dass die MT eines LOT-Grundterms entweder gelernt oder angeboren sein muss, und der Aussage, dass |cat|s MT nicht angeboren ist, auf die Aussage, dass |cat|s MT gelernt werden muss.

Rupert kritisiert diesen Argumentationsschritt, indem er nachzuweisen versucht, dass der Begriff von „angeboren“, welcher Cummins erlaubt |cat|s MT als nicht angeboren zu bezeichnen, weniger MTs umfasst als der Begriff „angeboren“ in der Disjunktionsbehauptung. Dies geschieht, indem Rupert nachzuweisen versucht, dass es MTs gibt, welche weder angeboren im starken Sinne noch gelernt sind.

Als ein Ergebnis seiner Argumente nimmt Rupert in Anspruch gezeigt zu haben, dass Cummins keine Begründung dafür gibt, dass der Begriff „angeboren“ in der Disjunktionssatzung extensionsgleich ist mit dem Begriff „angeboren“, der |cat|s MT ausschließt.

„If I have accurately captured Cummins`s thinking, however, he errs in his treatment of [...] primitives: he gives no argument to show that a LOT term`s being unlearned, [...], requires that the newborn have (P2)-knowledge about that to which the term applies.”<sup>30</sup> (Rupert 2001, S. 515).

Hieraus folgert Rupert, dass es durchaus LOT-Grundterme gibt, deren MTs auf andere Weise entstehen als gelernt zu werden oder bei der Geburt schon vorhanden zu sein:

„The causal theorist should feel free to account for the fixation of these terms` contents by appeal to something other than the processes allowed by (P1).”<sup>31</sup> (Rupert 2001, S. 521).

Ob |cat|s MT nun im schwachen Sinne angeboren ist oder gelernt werden muss, lässt Rupert hier allerdings offen.

„Nonetheless, what has been said so far does not show that any particular LOT term, such as `cat`, is unlearned; nor does it show that all LOT primitives are unlearned.”<sup>32</sup> (Rupert 2001, S. 521 f.).

---

<sup>30</sup> Rupert geht in seiner Argumentation nach und nach von „mediating knowledge being unlearned“ zu der missverständlichen Ausdrucksweise „LOT-terms` being unlearned“ über. Im Kontext wird klar, dass hier nichts über den LOT-Term ausgesagt werden soll, sondern etwas über seine MT (siehe Fußnote 16).

Mit „knowledge at birth“ ist hier eine im starken Sinne angeborne MT gemeint (siehe Fußnote 27).

<sup>31</sup> Die Prozesse der Entstehung von MT`s, die in der Disjunktionssatzung (P1) genannt werden, sind Lernen und vor der Geburt Entstehen (also angeboren sein).

<sup>32</sup> Hier ist mit |cat| ebenfalls |cat|s MT gemeint (siehe Fußnote 30).

### 2.3.6. Kommentar

Rupert verfolgt mit seiner Kritik an Cummins' erstem Argumentationsschritt das Ziel Cummins' speziellen Schluss für die MT von |cat| in Frage zu stellen. Obwohl Ruperts Einwände, wie ich in 2.3.6. (A) bis 2.3.6. (F) darstellen werde, stellenweise problematisch sind, bringt er genug Argumente vor um Cummins' Schluss für |cat| als zumindest unbegründet zu erweisen (siehe 2.3.6. (F)). Da |cat| bei Cummins als paradigmatisches Beispiel für eine große Menge von MTs fungiert, ist sein Argument durch Ruperts Kritik dahingehend geschwächt, dass diese große Menge von MTs, wie Rupert zeigt, nicht mehr als notwendigerweise gelernt angesehen werden muss.

#### 2.3.6. (A) Plastizität funktionaler Architektur

Ruperts Kritik an der Aussage, die funktionale Architektur eines repräsentationalen Systems wäre unveränderbar, welche Cummins fälschlich Pylyshyn zuschreibt und als Charakteristikum von CT und LOTH ansieht, ist berechtigt.

Erstens weist Pylyshyn an einigen Stellen in Pylyshyn (1984) sogar explizit darauf hin, dass die funktionale Architektur veränderbar ist, z.B.: „Some of these mechanisms [der funktionalen Architektur] may be quite rigid, whereas others may be alterable in various ways by the environment.“ (Pylyshyn 1984, S. 30).

Zweitens sind die von Rupert angeführten Beispiele von Veränderungen neuronaler Architektur sowohl etablierte Forschungsgegenstände der Neurowissenschaften als auch Vorgänge, welche größtenteils nicht als explizites Lernen beschrieben werden können.

Jedoch bezieht sich Ruperts Argumentation hier stets auf funktionale Architektur allgemein und nicht auf funktionale Architektur, welche MTs implizit repräsentiert.

Im Vergleich zu dem argumentativen Ziel, welches Rupert für diesen Teil seiner Kritik vorgibt, ist sein Angriff auf Cummins Argument somit sehr schwach.

Eigentlich geht es Rupert um die rekonstruierte Behauptung von Cummins, dass in der Architektur enthaltene MTs im starken Sinne „angeboren“, also nicht veränderbar, sind.

Diese Aussage wird von Rupert jedoch nur sehr indirekt kritisiert, indem er anstatt der Aussage selbst lediglich eine mögliche Begründung der Aussage widerlegt.

Cummins scheint die Behauptung, dass MTs, die implizit in der Architektur realisiert sind, sich nicht verändern, also im starken Sinne „angeboren“ sind, damit zu begründen, dass sich funktionale Architektur nach CT und LOTH insgesamt nicht verändert.

Diese Begründung der Behauptung wird bei Rupert widerlegt. Die Behauptung selbst wird jedoch durch seine Argumentation nicht direkt angegriffen.

Eine direkte Kritik an der eben genannten Behauptung wäre eine Begründung dafür, dass durch die Veränderungen der neuronalen Architektur nicht nur irgendwelche Strukturen entstehen können, sondern tatsächlich MTs von LOT-Grundtermen.

Dies scheint Rupert rückblickend auch für seine Kritik in Anspruch zu nehmen. Das Ergebnis der von mir in 2.3.5. (A) dargestellten Argumentation fasst Rupert so zusammen: „[...] architecture can change postnatally in ways that do not consist in learning but nonetheless result in the subject's holding new implicit theories.“ (Rupert 2001, S. 507). Dieses Ergebnis wird jedoch durch keinen Teil seiner Argumentation nahe gelegt.

Dass funktionale Architektur überhaupt veränderbar ist, wird von Rupert durchaus schlüssig begründet. Ob diese Veränderungen auch für Architektur möglich sind, welche eine MT ge-

speichert hat, wird von Rupert genauso wenig diskutiert, wie die Frage, ob und wie aus solchen Veränderungen eine neue MT hervorgehen könnte.

Als nachvollziehbares Ergebnis der in 2.3.5. (A) dargestellten Argumentation ist also nur festzuhalten, dass Cummins' Behauptung, in der Architektur implizite MTs wären im starken Sinne angeboren, unzulänglich begründet ist.

Keine Gründe hat Rupert jedoch dafür angegeben, dass die Behauptung falsch oder zumindest unplausibel wäre, dass also in der Architektur implizite MTs tatsächlich nach der Geburt entstehen können.

#### 2.3.6. (B) Funktionale Architektur und implizite MTs

Rupert ist bezüglich seiner Bemerkung, dass Cummins seine Implikationsvermutung nicht begründet, Recht zu geben.

Jedoch ist fraglich, welchen Zweck die Kritik an der Implikationsvermutung im Hinblick auf Ruperts argumentatives Ziel, die Unterschiedlichkeit der Bedeutungen von „angeboren“ in (P1) und (P2) zu zeigen, überhaupt erfüllen soll.

Erstens handelt es sich schließlich nur um eine Vermutung, welche weder Cummins' Disjunktionsbehauptung oder (P2) irgendwie stützt noch von Cummins für irgend ein anderes Teilargument explizit als Begründung benutzt wird. „If it is innate, it might [!], in some sense [!], be implicit in the architecture.“ (Cummins 1997, S. 536, Fußnote 5).

Zweitens ist nicht klar, wie die Kritik an der Implikationsvermutung, eigentlich Ruperts Behauptung, der Begriff „angeboren“ in (P1) und der Begriff „angeboren“ in (P2) wären nicht extensionsgleich, stützen soll. Durch die in 2.3.5. (A) beschriebenen Argumente meint Rupert schließlich, bereits gezeigt zu haben, dass implizit in der Architektur realisierte MTs nicht unbedingt im starken Sinne „angeboren“ sein müssen.

Eine mögliche Interpretation wäre, dass Rupert die Existenz einer weiteren Klasse von MTs zeigen möchte, die im schwachen Sinne angeboren sind, im starken jedoch nicht. Allerdings führt Rupert auch später weder Beispiele für MTs, die implizit aber nicht in der Architektur realisiert sind,<sup>33</sup> an noch skizziert er allgemein, wie MTs, die nicht in der Architektur realisiert sind, implizit sein können.

Unter dieser Interpretation haftet der in 2.3.5. (B) beschriebenen Kritik die selbe Schwäche an, wie den Argumenten aus 2.3.5. (A). Es wird mit einigem rhetorischen Aufwand gezeigt, dass Cummins' Voraussetzungen nicht begründet sind. Jedoch kann Ruperts Kritik Cummins' Argumente darüber hinaus weder widerlegen noch greift sie diese überhaupt irgendwie direkt an.

#### 2.3.6. (C) „angeboren“-Definitionen bei Rupert

Ruperts argumentatives Ziel in seiner Kritik an Cummins' erstem Argumentationsschritt ist es zu zeigen, dass der Begriff „angeboren“ im Sinne von (P2) nicht extensionsgleich ist mit dem Begriff „angeboren“ in (P1). Er muss also plausibel machen, dass es MTs gibt, die im schwachen Sinne „angeboren“ sind, im starken Sinne jedoch nicht.

Rupert versucht nun die Existenz solcher MTs als wahrscheinlicher erscheinen zu lassen, indem er Definitionen von „angeboren“ referiert, welche den starken Gebrauch in (P2) erweitern.

---

<sup>33</sup> Ein solches Beispiel wäre die kodierte MT, welche ich am Anfang von 2.3.2. (B) 3. beschrieben habe.



Dieses Vorgehen ist aus drei Gründen problematisch:

Erstens beziehen sich diese Definitionen zum größten Teil ganz allgemein auf Merkmale (auch kognitive) von Organismen und nicht speziell auf MTs von LOT-Grundtermen. Dass eine schwache Definition des Begriffs „angeboren“ sich für die Erklärung der Entstehung beliebiger Merkmale von Organismen als nützlich erwiesen hat, ist kein Anhaltspunkt dafür, dass dies auch für MTs von LOT-Grundtermen gilt, geschweige denn dafür, dass tatsächlich MTs existieren, die in einem schwachen Sinne aber nicht im starken Sinne „angeboren“ sind.

Zweitens müsste, damit die Benutzung der Definition wirklich ein Indiz für die Existenz von im schwachen aber nicht im starken Sinne angeborenen MTs ist, sichergestellt sein, dass die Definition nicht etwa noch schwächer ist, als die Minimaldefinition von „angeboren“ als nicht gelernt (in (P1)).

Falls nämlich die Definition auch auf gelernte MTs zuträfe, wäre die Benutzung dieser Definition kein Indiz für die Existenz von MTs, die nicht im starken Sinne angeboren sind, jedoch auch nicht gelernt sind. In diesem Fall könnte die Extension von „angeboren“ im starken (P2) Sinne (bei der Geburt vorhanden) der Extension von „angeboren“ im schwachen (P1) Sinne (nicht gelernt) entsprechen, aber durchaus kleiner sein als die Extension von „angeboren“ im Sinne der jeweiligen Definition, da diese schließlich gelernte Terme nicht ausschließt. Zwei der vier Definitionen lassen jedenfalls keine eindeutige Abgrenzung gegen gelernte MTs erkennen.

Drittens kann diese Aufzählung von „angeboren“-Definitionen, wie auch die Erörterung in 2.3.5. (B), nicht als direktes Gegenargument gegen Cummins ersten Argumentationsschritt angesehen werden, sondern zeigt nur, dass Cummins Voraussetzungen nicht zureichend begründet sind.

Für die tatsächliche Existenz von MTs, die weder gelernt noch im starken Sinne angeboren sind, wird bei Rupert nicht argumentiert. Er kann mit der Aufzählung der „angeboren“-Begriffe nur zeigen, dass Cummins die Eigenschaft bei der Geburt vorhanden zu sein für MTs nicht schon deshalb mit der Eigenschaft nicht gelernt zu sein gleichsetzen muss, weil es keine Definitionen von „angeboren“ gäbe, die MTs zuließen, welche bei der Geburt nicht vorhanden sind, jedoch auch nicht gelernt sind.

### 2.3.6. (D) Die Definitionen

Zur ersten Definition von J. L. Elman et al.:

Diese Definition von „angeboren“ enthebt bei der Entscheidung, ob etwas angeboren ist oder nicht, zwar von dem Problem zu beurteilen, was bei der Geburt implizit oder explizit „vorhanden“ ist. Jedoch wird die Definition durch den Bezug auf „normal developmental experiences“ sehr vage. So ist die Wahrscheinlichkeit, dass jemand die MT für |cat| entwickelt, vermutlich stark davon beeinflusst, ob die beurteilungsrelevanten „normalen Entwicklungsbedingungen“ stets die Konfrontation mit Katzen enthalten.

Zuzustimmen ist Rupert jedenfalls dahingehend, dass diese Definition viele Merkmale einschließt, die nicht im starken (P2) Sinne angeboren sind.

Rupert hält die Benutzung der ersten Definition von „angeboren“ bei J. L. Elman et al.(1996) in den Kognitionswissenschaften für ein Indiz für die Richtigkeit seiner Kritik an Cummins' erstem Argumentationsschritt, da die Definition nicht erfordert, dass ein angeborenes Merkmal schon bei der Geburt vorhanden ist.

„[...] this definition does not require that (P2)-knowledge be present in the newborn in order for some later-emerging behaviour or cognitive structure to qualify as innate.” (Rupert 2001, S. 512).

Alle drei in 2.3.6. (C) genannten Kritikpunkte treffen bezüglich dieser Definition auf Ruperts Argumentation zu:

Erstens kann man daraus, dass diese Definition für beliebige Merkmale von Organismen brauchbar ist, nicht darauf schließen, dass sie auch für MTs von LOT-Grundtermen sinnvoll ist.

Zweitens könnte es unter gewissen „normalen Entwicklungsbedingungen“ auch sehr wahrscheinlich sein, dass MTs von gewissen LOT-Grundtermen explizit gelernt werden.

Drittens gibt die Tatsache, dass diese Definition auch für MTs von LOT-Grundtermen benutzt werden könnte, keinen Anhaltspunkt dafür, dass es tatsächlich MTs gibt, die der Definition entsprechen, aber nicht bei der Geburt vorhanden sind.

Zur zweiten Definition von J. L. Elman et al.:

Bei dieser Definition treten zwei der drei oben genannten Schwächen der argumentativen Wirkung noch deutlicher zutage.

Rupert zählt diese Definition von „angeboren“ zu denen, die schwächer sind als die Definition, welche von (P2) gefordert wird. „Here again, innateness does not require (P2)-knowledge.“ (Rupert 2001, S. 512).

Für allgemeine Merkmale von Organismen ist die Definition von „angeboren“ tatsächlich eindeutig schwächer als die Definition von „angeboren“ im Sinne von „bei der Geburt vorhanden“.

Ob dies allerdings auch für MTs von LOT-Grundtermen gilt, ist fraglich.

In einem wesentlichen Punkt jedenfalls geht diese Definition nicht über die starke, von (P2) geforderte Definition hinaus.

Die starke Definition von „angeboren“ wird bei Rupert auch so beschrieben:

„Let me call (P2) *-knowledge* the kind of knowledge Cummins has in mind when he dismisses out of hand the possibility that newborns know enough about cats to facilitate the fixation of the content of ‘cat’ per CT.“ (Rupert 2001, S. 511 f.).

Die wesentliche Eigenschaft der starken Definition ist also, die MT von |cat| auszuschließen.

Dieses könnte die vorliegende Definition aber auch erfüllen. Jedenfalls scheint es nicht plausibel, dass ein Organismus die Fähigkeit, Katzen zu detektieren, falls er diese Fähigkeit nicht von Geburt an hat, erwerben kann, ohne, dass dieser Erwerb teilweise durch Ursachen außerhalb des Organismus erklärt werden müsste.

Die Definition ist also im für die Diskussion entscheidenden Punkt nicht einmal schwächer, als die von (P2) geforderte Definition.

Lernen von MTs wird durch diese Definition zumindest ausgeschlossen<sup>34</sup>.

Allerdings kann die Tatsache, dass diese Definition des Begriffs „angeboren“ in der Psychologie benutzt wird, nicht als Anhaltspunkt dafür dienen, dass es überhaupt LOT-Grundtermen gibt, deren MTs bei der Geburt zwar noch nicht vorhanden waren, jedoch durch ausschließlich dem Organismus interne Ursachen erworben wurden.

Zur Definition von R. Samuels:

Die Definition von R. Samuels stellt eine Verschärfung der ersten Definition von Elman et al. dar. Die Forderung, welche ein Merkmal erfüllen muss um „angeboren“ zu sein, wird hier erstens vom wahrscheinlichen zum sicheren Auftreten in allen Standardumwelten verschärft.

---

<sup>34</sup> Es wäre zwar durchaus mit CT und LOTH vereinbar, dass MTs von LOT-Grundtermen, deren Gehalte in Eigenschaften bestehen, welche Zuständen von körperinneren Organen zugeschrieben werden, gelernt (sogar explizit) werden könnten, ohne dass dazu Ursachen außerhalb des Organismus notwendig wären. Die Diskussion bezieht sich hier jedoch ausschließlich auf „distal properties“. Für andere ist gemäß 2.2.1. nicht einmal Transduktion ausgeschlossen.

Zweitens wird hier der Träger des betreffenden Merkmals<sup>35</sup> als Gruppe genidentischer Organismen genauer bestimmt.

Die Probleme der ersten Definition von J. L. Elman lassen sich bei der Definition von R. Samuels analog feststellen.

Es ist weder klar, ob die Definition im schwachen Sinne angeborene MTs ausschließt, noch bezieht sich die Definition auf MTs direkt, gibt also auch keinen Anhaltspunkt für die Existenz einer bestimmten Klasse von MTs.

Zur Definition von J. Fodor:

Die Formulierung von J. Fodor entspricht in der Tat der schwachen Bedeutung von „angeboren“, die Cummins für seine Disjunktionsbehauptung voraussetzen muss.

Somit kann die Existenz dieser Definition natürlich nicht die argumentative Rolle spielen, die Rupert im Voraus für seine Aufzählung von „angeboren“-Definitionen in Anspruch nimmt, nämlich als Indiz für die Falschheit von entweder (P1) oder (P2) zu dienen.

In diesem Zusammenhang ist Definition von „angeboren“ eher als ein Hinweis dafür zu sehen, dass Fodors Theorie mentaler Repräsentationen tatsächlich durch Cummins Kritik getroffen wird.

### 2.3.6. (E) Entwicklungspsychologische Paradigmen

Aus drei Gründen ist Ruperts Schluss von der Debatte zwischen selektionistischem und konstruktivistischem Ansatz auf die Existenz von MTs von LOT-Grundtermen, welche nicht gelernt und auch nicht angeboren sind, in Zweifel zu ziehen.

Erstens diskutiert Rupert in diesem Abschnitt nicht die Entstehung von MTs, sondern die Entstehung von physischen Trägern von LOT-Grundtermen.

Zweitens hat die referierte Debatte keinerlei direkten Bezug auf LOT-Grundterme oder deren MTs.

Und drittens sind die beschriebenen Entstehungsprozesse durchaus mit Cummins' Identifikation von nicht gelernten mit im starken Sinne angeborenen MTs verträglich, da die Entstehungsprozesse auch als Lernen beschrieben werden können.

In diesem Kapitel werde ich diese drei Einwände näher erläutern.

Zuerst zum Bezug auf LOT-Grundterme:

Neben dem Zweck einer Vorüberlegung zur Kritik an Cummins' zweitem Argumentationsschritt soll die Darstellung der Debatte zwischen konstruktivistischem und selektionistischem Ansatz folgende Funktion erfüllen: „[...] to underscore the false nature of Cummins' s dilemma offered in (P1), by describing some mechanisms likely to govern the emergence of nonsemantically individuated LOT terms in humans, mechanisms that would seem operative in cases of at least some primitive, unlearned LOT terms; [...].“ (Rupert 2001, S. 520).

Das Dilemma, welches Cummins in (P1) beschreibt, bezieht sich jedoch nicht auf LOT-Grundterme, sondern auf deren MTs. Cummins behauptet dort, dass MTs von LOT-Grundtermen entweder im starken Sinne angeboren oder aber gelernt sind.

---

<sup>35</sup> In der Darstellung der ersten Definition von Elman et al. lässt es Rupert offen, ob eine Aussage über ein „token“- Merkmal eines Individuums anhand möglicher Standard-Umwelten gemacht werden soll, oder über ein „typ“- Merkmal in einer Gruppe von Individuen anhand deren tatsächlicher Standard - Umwelten. Für die Diskussion von Cummins' erstem Argumentationsschritt ist die Intention der Definition diesbezüglich unerheblich, da die Definition, sofern sie den allgemeinen in 2.3.6. (C) genannten Problemen bezüglich der argumentativen Wirkung entgegen würde, sowohl in der einen als auch in der anderen Interpretation, Cummins' ersten Argumentationsschritt gleichermaßen treffen würde.

Es erscheint deshalb merkwürdig, dass Rupert seine Diskussion der eben genannten entwicklungspsychologischen Ansätze auf LOT-Symbole<sup>36</sup> beschränkt.

Von der Herkunft eines LOT-Symbols kann nämlich kein Schluss auf die Herkunft der MT des entsprechenden LOT-Grundterms gezogen werden.

Gemäß LOTH und CT fungieren die LOT-Symbole tatsächlich nur als Symbole, deren Struktur oder Herkunft nicht notwendigerweise im Zusammenhang mit ihrem Gehalt stehen.

Die allgemeine Formulierung der CT und LOTH lässt es zu, dass ein gelerntes LOT-Symbol durch ein im starken Sinne angeborenes LOT-Symbol ersetzt wird, ohne dass sich der Gehalt oder die MT, welche ja den Gehalt bestimmt, ändern müsste. Auch der Status der MT als gelernt oder angeboren müsste sich bei solch einer Vertauschung nicht ändern.

Wenn man lediglich die allgemeine Version der CT und LOTH voraussetzt, ist also kein Schluss von den Modalitäten der Entstehung von LOT-Symbolen auf die Richtigkeit von Behauptungen über den Inhalt von LOT-Termen oder MTs, wie sie Cummins in (P1) macht, möglich.

Um eine Beziehung zwischen der Entstehung von LOT-Symbolen und der Entstehung von MTs von LOT-Grundtermen herzustellen, wären über die Grundversionen von LOTH und CT hinausgehende Zusatzannahmen bezüglich eines Zusammenhangs zwischen der Entstehung von Symbolen und der Entstehung von MTs notwendig, wie Rupert sie am Rande vorschlägt:

„In many cases, the very causal interactions with the environment that shape the brain's non-semantically characterized representational resources at the same time contribute significantly to the fixation of the content of the units so shaped, [...]“ (Rupert2001, S. 521, Fußnote 38).

Um Cummins' Voraussetzung (P1) anzugreifen, kann Rupert jedoch nur die Grundversionen von LOTH und CT voraussetzen. Dies nimmt er dann auch für seine Argumentation in Anspruch:

„[...] the arguments of the preceding and following sections [alle in der vorliegenden Arbeit besprochenen Argumente Ruperts] stand independent of any controversial claims about the relation between cognitive activity and development and their neurological bases.“ (Rupert 2001, S. 521, Fußnote 37).

Wenn man auf solche Zusatzannahmen verzichtet, kann Ruperts Beschreibung der obengenannten entwicklungspsychologischen Debatte aber nicht als Argument gegen Cummins' ersten Argumentationsschritt angesehen werden.

Zum zweiten Einwand:

Obwohl sich Ruperts Kritik nun durch die Beschränkung auf LOT-Symbole bereits als unwirksam herausgestellt hat, scheint es sinnvoll, zu prüfen, ob Ruperts Diskussion der entwicklungspsychologischen Ansätze als ein Argument gegen Cummins' ersten Argumentationsschritt gelten könnte, wenn man von dem eben festgestellten Problem absieht und die Argumentation direkt auf MTs von LOT-Grundtermen bezieht.

Ich werde hier also noch untersuchen, ob das Argument mit Bezug auf MTs als Argument gegen Cummins' ersten Argumentationsschritt gelten kann.

---

<sup>36</sup> Ein LOT-Grundterm kann sowohl semantisch als auch biologisch und syntaktisch beschrieben werden. Rupert räumt jedoch in seiner Kritik des zweiten Argumentationsschrittes ein, dass es LOT-Grundterme ohne semantische Beschreibung gibt. Er bezeichnet die biologisch und syntaktisch beschriebenen LOT-Grundterme als „non-semantically individuated LOT terms“. Bei Cummins und anderen werden Repräsentationen, wenn lediglich auf ihre syntaktischen und biologischen Eigenschaften Bezug genommen wird, als „vehicle“ des Gehaltes bezeichnet.

Ich bezeichne LOT-Terme bei ausschließlicher Bezug auf die syntaktischen und biologischen Eigenschaften hier als „Träger der Repräsentation“ oder bei Voraussetzung der LOTH als „LOT-Symbole“.

Ob es sich bei LOT-Symbolen ohne semantische Repräsentation überhaupt noch um LOT-Terme handelt, werden ich in 2.4.4. besprechen.

Dass es überhaupt möglich ist Ruperts Argument ohne weiteres auf MTs anzuwenden, verweist schon auf ein weiteres Problem.

Es liegt ein ähnlicher Einwand nahe, wie ich ihn schon im Kapitel 2.3.6. (C) und (D) mehrfach formuliert habe.

Rupert schließt hier aus einer Voraussetzung einer Debatte, die allgemein über die Entstehung funktionaler Hirnstrukturen geführt wird, darauf, dass diese Voraussetzung auch für die Entstehung von MTs (bzw. von LOT-Symbolen) gilt.

Die von Rupert im Zusammenhang der Selektionismus/Konstruktivismus-Debatte angeführten Quellen argumentieren jedoch nicht mit direktem Bezug auf MTs von LOT-Grundtermen oder LOT-Symbole.

Es geht dabei lediglich allgemein um die Entstehung funktionaler Hirnstrukturen.

So führt Rupert die Selektionismus/ Konstruktivismus-Debatte als Versuch ein folgende

Frage zu beantworten:

„How does experience contribute to the development of the functional properties of various [!] neural structures?“ (Rupert 2001, S. 518).

Da die funktionalen Strukturen, deren Entstehung in den empirischen Neurowissenschaften derzeit untersucht werden, dort zum allergrößten Teil nicht als LOT-Symbole aufgefasst werden, ist es sehr fraglich, ob man aus solchen Debatten irgendetwas über LOT-Symbole oder MTs schließen kann.

Jedenfalls gäbe es auch im Falle, dass MTs von LOT-Grundtermen entweder angeboren im starken Sinne oder gelernt sein müssten, genügend Anlass die von Rupert skizzierte Debatte zu führen.

Somit ist die Tatsache, dass in der Selektionismus/Konstruktivismus Debatte eine bestimmte Voraussetzung nicht hinterfragt wird, nicht einmal als ein Indiz für die Falschheit von Cummins' Voraussetzung (P1) anzusehen.

Des Weiteren sei hier, unabhängig von der inhaltlichen Wirkungslosigkeit, auf die befremdliche rhetorische Struktur in Ruperts Argument hingewiesen:

Rupert möchte darlegen, dass postnatale Entwicklung funktionaler neuronaler Strukturen durch Interaktion mit der Umwelt, welche nicht unbedingt als Lernen beschrieben werden kann, in den Neurowissenschaften unumstritten ist.

Ausschließlich um dies zu begründen referiert Rupert ausführlich eine in den Neurowissenschaften geführte Debatte, welche sich überhaupt nicht direkt auf diese Tatsache bezieht, sondern diese voraussetzt.

Für die direkte Begründung dieser Tatsache begnügt sich Rupert mit einem ergänzenden Einschub während der Einführung der Selektionismus/Konstruktivismus-Debatte: „[...] – the experimental evidence seems [!] to have settled that issue – [...].“ (Rupert 2001, S. 518; für das ganze Zitat siehe 2.3.5 (D)).

Dies erscheint insbesondere insofern merkwürdig, als dass sogar in aktuellen neurowissenschaftlichen Lehrbüchern zuhauf direkte Begründungen dieser Voraussetzung, inklusive der entsprechenden „experimental evidence“, zu finden sind.

In z.B. Dudel, J.; et al (1996, S. 485 ff.), Gazzaniga, M. S. et al. (2002, S. 611 ff.) oder Kolb, B. (1996, S. 420 ff.) sind sowohl direkte Formulierungen als auch eindeutige experimentelle Belege für die Aussage enthalten, die Rupert eigentlich hervorheben möchte: „[...], experience can modify the functional organization of the nervous system.“ (Gazzaniga, M. S. 2002, S. 614)

Es sind in der Geschichte der Wissenschaften mehrfach lebhaft Diskussionen anhand von Voraussetzungen, welche sich später als falsch herausgestellt haben, geführt worden.

Daher ist eine direkte Begründung für eine Behauptung, falls eine solche möglich ist, eigentlich einer indirekten Begründung durch die Gebräuchlichkeit der Behauptung als Voraussetzung in anderen Debatten, vorzuziehen.

Auch der dritte Einwand gegen Ruperts Argumentation wurde schon in ähnlicher Form gegen die Argumente aus 2.3.5. (C) und (D) vorgebracht.

Rupert nimmt die Aussage, dass es MTs<sup>37</sup> gibt, welche nicht im starken Sinne angeboren sind, also durch Erfahrung entstehen, als Argument gegen Cummins' Disjunktionsbehauptung.

Dies kann jedoch nur funktionieren, wenn gezeigt wird, dass diese MTs nicht als gelernt beschrieben werden können.

Dass die diskutierten Veränderungen in der funktionalen Hirnstruktur nicht grundsätzlich als Lernvorgang beschrieben werden können, dafür hat Rupert nicht einmal allgemein für beliebige funktionale Strukturen argumentiert.<sup>38</sup>

Dass funktionale Strukturen (insbesondere MTs) postnatal entstehen können, dagegen hat sich Cummins gar nicht ausgesprochen, sondern nur dafür, dass dieses Entstehen dann als Lernprozess beschrieben werden sollte.

Somit kann ein Beleg für die Existenz von nicht im starken Sinne angeborenen funktionalen Strukturen nicht als Argument gegen Cummins' Disjunktionsbehauptung gelten.

### 2.3.6. (F) Cummins' Schluss auf die Notwendigkeit des Lernens von |cat|s MT

Ruperts Kritik an Cummins' erstem Argumentationsschritt fällt in ihrer Wirkung, gemessen am argumentativen Aufwand, äußerst schwach aus.

Cummins' Ergebnis wird nicht einmal explizit angegriffen. Rupert nimmt als Ziel seiner Bemühungen dann auch nur in Anspruch gezeigt zu haben, dass Cummins seine Prämissen nicht begründet hat.

Anstatt die CT aktiv zu verteidigen, indem einige von Cummins' Prämissen, welche ihn dann darauf schließen lassen, dass |cat|s MT gelernt ist, durch Gegenargumente in Frage gestellt werden, konstruiert Rupert in drei Abschnitten merkwürdige potentielle Argumente für Cummins' Prämissen, welche dann gut begründet abgelehnt werden.

Für die in 2.3.5. (A) dargestellten Argumente sieht diese Strategie so aus:

Rupert zeigt, dass aus CT und LOTH nicht folgt, dass alle funktionalen Strukturen in einem repräsentationalen System entweder angeboren oder gelernt sind.

Somit könnte Cummins seine Aussage, dass aus CT und LOTH folgt, dass |cat|s MT entweder angeboren oder gelernt ist, nicht damit begründen, dass diese Disjunktion nach CT und LOTH grundsätzlich für alle neuronalen Strukturen gilt.

Hierdurch wird die Plausibilität von Cummins' Prämisse jedoch nicht wesentlich verringert, da die Aussage, dass alle neuronalen Strukturen entweder gelernt oder angeboren sind, so unsinnig ist, dass sie vermutlich weder Cummins selbst noch einem Rezipienten seiner Argumente als mögliche Begründung der Disjunktionsaussage in den Sinn käme.

Eine wirkungsvolle Verteidigung der CT wäre eine zumindest grobe Skizze, wie eine MT entstehen könnte ohne gelernt zu werden oder angeboren zu sein. Alle Beispiele für von Lernprozessen verschiedene Möglichkeiten der postnatalen Veränderung neuronaler Strukturen, die Rupert anspricht, kommen jedenfalls, wie z. B. Hirnverletzungen (Rupert 2001, S. 505), nicht für eine plausible Erklärung der regulären Entstehung von MTs infrage.

Ähnlich sind Ruperts andere Argumente zu beurteilen. Die in 2.3.5. (C) dargestellten Definitionen von „angeboren“ zeigen nur, dass Cummins MTs, die nicht gelernt und nicht angebo-

---

<sup>37</sup> Hier sei die Modifikation des Argumentes, welche ich oben infolge meines ersten Kritikpunktes vorgenommen habe, vorausgesetzt.

<sup>38</sup> In den Neurowissenschaften werden Strukturveränderungen, welche viel primitiver sind, als systematische Veränderungen von impliziten MTs für die Detektion von „distal properties“, bereits als „Lernen“ bezeichnet. So werden z.B. in den Arbeiten von Gilbert (1994) oder Schoups (2002) Veränderungen der neuronalen Architektur explizit als Lernen bezeichnet, wobei die durch die betreffende Architektur realisierten Wahrnehmungsprozesse nach Pylyshyns Schema, wenn nicht sogar als Transduktion, dann als nicht „kognitive penetrabel“ Prozesse, gelten würden.

ren sind, nicht schon deshalb ausschließen kann, weil in den empirischen Kognitionswissenschaften grundsätzlich keine Dinge – irgendwelcher Art – bekannt wären, die weder angeboren noch gelernt sind. Das würde jedoch auch kaum jemand bezweifeln.

Dieselbe Bemerkung trifft, wie ich in 2.3.6. (E) gezeigt habe, auch auf die Diskussion der entwicklungspsychologischen Paradigmen aus 2.3.5. (E) zu.

So gesteht Rupert auch zu, Cummins' ersten Argumentationsschritt zwar bezüglich dessen Begründung bezweifelt, jedoch bezüglich des Ergebnisses nicht widerlegt zu haben. „[...], what has been said so far does not show that any particular LOT term, such as 'cat', is unlearned; nor does it show that all LOT primitives are unlearned.” (Rupert 2001, S. 521 f.).

Somit kann Ruperts Kritik an Cummins' erstem Argumentationsschritt also nur als Hinweis darauf gelten, dass Cummins' Argumentationsschritt nicht näher begründet ist.

## 2.4. Lernen von LOT Termen

### 2.4.1. Cummins' zweiter Argumentationsschritt

In seinem zweiten Argumentationsschritt beschreibt Cummins einen Widerspruch zwischen dem Ergebnis des ersten Argumentationsschrittes und der CT sowie der NTDP.

Im ersten Argumentationsschritt begründet Cummins, dass die MT von |cat| explizit gelernt werden muss. Nach Cummins erfordert die LOTH, dass dafür explizite LOT-Sätze über |cat|s Gehalt formuliert werden.

„Learning the theory will require formulating and confirming hypotheses such as these:

(A) Cats have whiskers.

(B) Cats have four legs.

(C) Cats have fur.

According to LOT [hier im Sinne von LOTH], these hypotheses are represented as sentences in LOT.” (Cummins 1997, S. 537).

Dafür muss der LOT-Grundterm |cat| gebildet werden, was, gemäß der in 2.2.3. besprochenen Grundannahmen der CT, erfordert, dass |cat|s Gehalt richtig detektiert wird.

Für die Detektion des Gehaltes eines LOT-Grundterms muss gemäß der NTDP (siehe 2.2.1.) jedoch wiederum eine MT vorausgesetzt werden. Für |cat| folgt allerdings nach Cummins aus der LOTH, dass diese MT explizit gelernt werden muss (das Ergebnis des ersten Argumentationsschrittes).

„As such, they [die LOT-Sätze, welche die MT über |cat|s Gehalt bilden; s. o.] require for their formulation, a symbol for cats, that is, a |cat|. But, according to CT, you cannot have a |cat| until you have the ability to detect cats. According to psychology, you cannot have the ability to detect cats until you have a theory of cats. According to LOT, you cannot have a theory of cats until you have |cat|s.” (Cummins 1997, S. 537).

Hieraus schließt Cummins, dass die LOTH und die CT nicht zusammen mit der NTDP wahr sein können.

„So, you cannot make the conjunction of LOT and CT compatible with psychology.”<sup>39</sup> (Cummins 1997, S. 537).

Das bedeutet, dass mindestens eine der Grundannahmen – LOTH, CT oder NTDP – falsch ist. Da Cummins zufolge erstens die NTDP eine grundsätzlich wahre Tatsache ist, und zweitens aufgrund der in 1.2.1. dargestellten Implikation die Falschheit der LOTH auch die Falschheit

---

<sup>39</sup> LOT im Sinne von LOTH (siehe Fußnote 5).

der CT impliziert, folgert er aus der Falschheit der Konjunktion von LOTH, CT und NTDP, dass die CT falsch ist<sup>40</sup>.

### **2.4.2. Kommentar**

Cummins setzt hier drei Implikationsbeziehungen voraus und zeigt, wie so aus CT, NTDP und LOTH eine offenbar falsche Aussage folgt:

Die Voraussetzungen:

- (1) Aus LOTH folgt, dass Lernen einer MT explizites Lernen (im Sinne von 2.3.2. (B) 1.) bedeutet und damit die Fähigkeit zur Bildung des entsprechenden LOT-Grundterms erfordert.
- (2) Aus CT folgt, dass es zur Bildung eines LOT-Grundterms notwendig ist, dass dessen Gehalt detektiert werden kann.
- (3) Aus NTDP folgt, dass für die Detektion eines LOT-Grundterms eine dem Term entsprechende MT vorhanden sein muss.

Hieraus folgt offenbar, dass, falls die MT eines LOT-Grundterms gelernt werden müsste, diese MT für den Lernvorgang schon vorhanden sein müsste.

Wenn man Lernen intuitiv plausiblerweise so versteht, dass das zu Lernende vor einem zeitlich begrenzten<sup>41</sup> Lernprozess nicht vorhanden ist und erst mit Beendigung des Lernprozesses vollständig vorhanden ist, dann ist eine solche Aussage, die behauptet, dass das zu Lernende – die MT – schon während des Lernprozesses oder vorher vorhanden sein muss, falsch.

Zusammen mit dem Ergebnis des ersten Argumentationsschrittes (dass es tatsächlich mindestens eine MT, nämlich die für |cat|, gibt, welche gelernt werden muss) implizieren die drei Voraussetzungen also wirklich eine offenbar falsche Aussage.

Um die Möglichkeit einer CT mentaler Repräsentation zu verteidigen, müsste man also entweder eine der drei Voraussetzungen oder das Ergebnis des ersten Argumentationsschrittes bezweifeln.

Die Gültigkeit der ersten Voraussetzung habe ich bereits in 2.3.2. (B) widerlegt.

Die zweite Voraussetzung ist schwer zu bezweifeln, da schließlich gemäß der verwendeten Definition der CT (siehe 2.2.3.) mittels einer kausalen Beziehung zwischen Repräsentiertem und Repräsentation (die als Detektion beschrieben werden kann) der Gehalt der Repräsentation definiert wird.

Auch eine Kritik der dritten Voraussetzung scheint aufgrund der minimalen Definition von MT als „some knowledge about ...“ (siehe 2.2.1.) nicht sehr vielversprechend.

Das Ergebnis des ersten Argumentationsschrittes ist von mir bereits durch Kritik an der Implikationsbehauptung widerlegt worden (siehe 2.3.2. (F)).

Obwohl Cummins' zweiter Argumentationsschritt in sich schlüssig ist, kann er nicht als Argument gegen die CT gelten, da ich sowohl die Voraussetzung (1) als auch das Ergebnis des ersten Argumentationsschrittes widerlegt habe.

### **2.4.3. Ruperts „coining terms“**

In seiner Kritik an Cummins' erstem Argumentationsschritt hat Rupert nur gezeigt, dass Cummins seinen Schluss von dem Nichtvorhandensein der MT für |cat| bei der Geburt auf die

---

<sup>40</sup> Aus  $(CT \rightarrow LOTH)$ , NTDP und  $\neg(CT \wedge LOTH \wedge NTDP)$  folgt logisch  $\neg CT$ .

<sup>41</sup> Cummins und Rupert gehen beide davon aus, dass repräsentationales Lernen für jede Repräsentation ein zeitlich begrenzter Prozess ist, da eine funktionsfähige Repräsentation erst nach Abschluss des Lernvorgangs richtig gebildet werden kann. „As the theory [|cat|s MT] gets elaborated, eventually [!], we are able to recognize cats reliably.“ (Cummins 1997, S. 538).



Notwendigkeit des expliziten Lernens von |cat|s MT nicht plausibel begründen kann. Aber er hat nicht zeigen können, dass es keine explizit gelernten MTs gibt, was Cummins' Argument schon durch Kritik am ersten Argumentationsschritt zu Fall gebracht hätte. Er gesteht sogar ausdrücklich zu, dass es explizit repräsentierte und explizit gelernte (was ja für Rupert dasselbe ist; siehe 2.3.2. (B) und 2.3.3.) MTs gibt (siehe 2.3.5. (E) und 2.3.6. (E)) und dass |cat|s MT dazu gehören könnte.

Somit muss Rupert für eine sinnvolle Verteidigung der CT auch Cummins' zweiten Argumentationsschritt kritisieren: „[...] I must now address the difficult case, where explicitly represented theory seems necessary to mediate content-fixing reliable detection.“ (Rupert 2001, S. 522).

Um Cummins' zweiten Argumentationsschritt zu kritisieren, versucht Rupert, einen Weg zu skizzieren, auf welchem die MT von |cat| explizit gelernt werden könnte ohne von dem Widerspruch getroffen zu werden, den Cummins für explizit gelernte MTs konstruiert hat.

„Assume, then, plausibly enough, that the explicit formulation of theory mediates reliable detection for 'cat', via such theoretical axioms as Cummins lists: '(A) Cats have whiskers. (B) Cats have four legs. (C) Cats have fur.' Is there a way for the subject to frame such explicitly represented axioms without requiring [...] that the subject already possess a concept with the content *cat* ? This is possible, [...].“<sup>42</sup> (Rupert 2001, S. 522).

Dieser Weg sieht so aus, dass man den physischen Träger des Symbols, welches später den Gehalt, eine Katze zu sein, haben soll, zuerst nicht als Repräsentation auffasst sondern nur durch seine syntaktischen Eigenschaften beschreibt. „[...] we envision the initial theoretical axioms as containing the nonsemantically individuated term 'cat', as opposed to a LOT term with the content *cat*.“ (Rupert 2001, S. 522).

In seiner Kritik an Cummins' erstem Argumentationsschritt beschreibt Rupert, wie postnatal neue funktionale Strukturen im Gehirn entstehen können, ohne dass diese während der Entstehung als Repräsentationen bezeichnet werden müssten. Hieraus leitet Rupert die Möglichkeit ab, dass die physischen Träger von LOT-Grundtermen bei ihrer Entstehung zwar ihre syntaktischen Eigenschaften nicht aber ihre semantischen Eigenschaften, den intentionalen Gehalt, haben könnten. „[...] at least some of the time after birth, the human brain/ cognitive system generates, or 'coins', new LOT terms, individuated nonsemantically; [...]. It is possible, then, that the subject's early theory of catness include a newly coined LOT term lacking the content *cat*, perhaps lacking content altogether. If this picture is correct, then the initial theoretical axioms that include the nonsemantically individuated LOT term we call 'cat' would more properly be thought of in the following way: (A) *t* (a newly coined, nonsemantically individuated LOT term) have whiskers; (B) *t* have four legs; (C) *t* have fur. Cummins' fundamental objection to CT is thus dissolved.“ (Rupert 2001, S. 522 f.).

Die für das Lernen notwendige explizite MT über |cat|s Gehalt enthält also nach Ruperts Theorie statt des LOT-Grundterms |cat| nur das syntaktisch beschriebene Symbol, welches nach dem Lernvorgang zur Repräsentation mit der Eigenschaft eine Katze zu sein als Gehalt wird. Rupert ist der Auffassung, dass für diese Möglichkeit des expliziten Lernens von MTs nicht der Widerspruch gilt, welchen Cummins im zweiten Argumentationsschritt konstruiert hat. „Cummins fundamental objection to CT is thus dissolved.“ (Rupert 2001, S. 522).

Ruperts Theorie erscheint in einigen Punkten recht vage formuliert. Insbesondere bleiben in der bis hierher zitierten Form die Fragen offen, wie der Vorgang des „coining“ funktioniert, bzw. was diesen eigentlich auslöst und wie die von Rupert beschriebene Vorstufe der korrekten MT an einem expliziten Lernvorgang teilnehmen soll.

Diese Fragen werden von Rupert in Form von Repliken auf potentielle Einwände gegen seine Theorie teilweise aufgeklärt. In den Kapiteln 2.4.3. (A) und 2.4.3. (B) werde ich Ruperts ge-

---

<sup>42</sup> Ob Rupert hier tatsächlich einen Lernvorgang beschreiben möchte, wird in 2.4.4. diskutiert.

nauere Beschreibungen des „coining“ und des entsprechenden Lernvorgangs zusammenfassend darstellen.

#### 2.4.3. (A) Funktionsweise des „coining“

Das von Rupert beschriebene „coining“ ist hier offenbar als Bildung eines LOT-Symbols<sup>43</sup> zu verstehen, welches dann durch einen Lernvorgang einen Gehalt bekommt und so zu einer LOT-Repräsentation wird. Rupert gibt in der Darstellung seiner Theorie genauere Auskünfte darüber, wie eine solche Symbolbildung funktioniert, und wie sie ausgelöst wird.

Für die Frage nach der Entstehung von LOT-Symbolen verweist Rupert zum einen auf seine Ausführungen zur Entstehung von LOT-Symbolen, welche ich in 2.3.5. (E) beschrieben habe, und nennt zum anderen Ansätze zur Erklärung der Entstehung stabiler funktionaler Einheiten von Neuronen aus Changeaux (1997) und Thelen (1994), welche Ruperts Meinung nach als physiologische Beschreibung von LOT-Symbolen dienen könnten.

Zur Frage, wie das „coining“ verursacht wird:

„What causes the brain/ cognitive system to produce a new LOT term that will be the subject of an explicit theory?“ (Rupert 2001, S. 524).

Diese Frage beantwortet Rupert folgendermaßen: „When the subject encounters a cluster of properties not sufficiently similar to any stored cluster (each cluster embodying a mediating theory), her cognitive system introduces a new primitive term.“ (Rupert 2001, S. 524).

Ausgelöst wird das „coining“ also durch einen Mechanismus, welcher wahrzunehmende „cluster“ von Eigenschaften dahingehend klassifiziert, ob sie einer im repräsentationalen System vorhandenen MT ähneln, wobei eine MT hier auch als gespeicherter „cluster“ von Eigenschaften aufgefasst wird (manchmal nennt Rupert eine als „cluster“ gespeicherte Theorie auch „prototype“).

Dieser Mechanismus wird von Rupert auch als Theorie beschrieben, welche eine Regel realisiert: „The Term-introduction rule could itself be either explicitly represented or implicit in the architecture.“ (Rupert 2001, S. 524).

„Whatever theory, implicit or otherwise, governs the introduction of new LOT terms may do so by measuring the match between a newly encountered item and stored, associated clusters of properties.“ (Rupert 2001, S. 525).

Das „coining“ eines LOT-Symbols erfordert also eine vorhergehende Messung der Ähnlichkeit zwischen einem „cluster“ von Eigenschaften und einer MT. Eine solche Messung stellt sich Rupert so vor:

„Here one might imagine a best-match, or ‘nearest neighbour’ system with a threshold of difference beyond which no match will be made, and instead a new term coined; [...]“ (Rupert 2001, S. 525, Fußnote 42).

Das „coining“ eines LOT-Symbols wird also durch einen Mechanismus im repräsentationalen System ausgelöst, welcher für wahrgenommene „cluster“ von Eigenschaftsinstanzierungen deren Ähnlichkeit zu im System schon vorhandenen MTs misst. Unterschreitet die Ähnlichkeit eine gewisse Schwelle, erfolgt das „coining“ eines neuen Symbols.

#### 2.4.3. (B) Lernen mit „coined terms“

Das „coining“ ist bei Rupert nur eine Vorbedingung des Lernens von MTs. Mit einem „coined term“ kann ohne eine Repräsentation eine Theorie über den Inhalt der Repräsentation formu-

---

<sup>43</sup> Siehe Fußnote 36.

liert werden, welche dann durch einen Lernvorgang verbessert werden kann. Wie ein solcher Lernvorgang aussehen könnte, darauf gibt Rupert auch einige Hinweise.

Für die einzelnen Lernschritte ist nach Rupert dieselbe Theorie zuständig, wie für das „coining“ selbst: „Implicit in the workings of the system is a theory that causes the coining of a new LOT term under some circumstance and the integration of new experiences and existing prototypes under others.” (Rupert 2001, S. 525). Die MT ist hier als „prototype“ aufzufassen.

Diese „Integration“, welche Rupert als Lernen bezeichnet, besteht in einer Veränderung der MT in Richtung des wahrgenommenen „cluster“ von Eigenschaften:

„[...] the system could be such that when a match is made and the stimulus differs from the prototype, the stimulus ‘pulls’ the prototype toward it in similarity space.” (Rupert 2001, S. 525, Fußnote 42).

In folgender Formulierung wird deutlicher, dass dies „pulling in similarity space“ tatsächlich auch als Veränderung einer expliziten LOT-MT beschrieben werden kann:

„The subject may later interact with a new item differing substantially from the item that caused the original coining of *t*, *without the new item’s differing enough from the original to cause the coining of a new LOT term*. The cognitive system responds by altering existing theoretical axioms or adding new axioms including *t*.” (Rupert 2001, S. 525).

Nachdem Rupert nun die Voraussetzung des Lernens („coining“) und den Lernvorgang selbst beschrieben hat, stellt sich die Frage, wie ein solcher Lernvorgang beendet wird, wie also die implizite das LOT-Symbol *t* enthaltende Theorie durch das Lernen zur richtig funktionierenden MT für |cat| wird. Am Ende des Lernprozesses soll *t* nach Rupert die Eigenschaft eine Katze zu sein als Gehalt haben und dann auch als |cat| die korrekte MT für die Detektion von |cat| konstituieren.

„[...] we should see [LOT-Symbole wie *t*] as potential representations, which have not yet, but may eventually, come into content-conferring relations to kinds or properties.” (Rupert 2001, S. 528).

Wie diese „potentiellen“ Repräsentationen dann zu tatsächlichen Repräsentationen werden können, darauf gibt Rupert erstaunlicherweise keine Antwort, sondern stellt die Beantwortung dieser Frage als Bedingung für eine konkretere Formulierung einer CT auf:

„This suggests a condition that must be met by a causal theory of content: it must assign content to newly coined, content-lacking LOT terms in a timely enough fashion that such terms possess content when our explanatory purposes demand that they do.“ (Rupert 2001, S. 527, Fußnote 47).

#### **2.4.4. Kommentar**

Rupert hat trotz seiner umfangreichen Kritik an Cummins’ Einteilung der MTs von LOT-Grundtermen (in im starken Sinne angeborene MTs und explizit gelernte MTs) nicht bezweifelt, dass es LOT-Grundterme gibt, deren MTs explizit gelernt werden müssen.

Diese Klasse von LOT-Grundtermen wird von Rupert und Cummins bei Geltung der CT und LOTH übereinstimmend beschrieben:

„Let us use learned terms to refer to those LOT terms for which, assuming CT and NTDP, it is a necessary condition of their acquiring content that the human subject formulate an explicit theory of that to which the terms apply; [...]“<sup>44</sup> (Rupert 2001, S. 502).

---

<sup>44</sup> Man könnte dieses Zitat auch so interpretieren, dass mit „acquiring content“ nicht der gesamte Lernprozess bezeichnet wird, dass Rupert also bloß gelernte explizite MTs meint, anstatt explizit gelernter expliziter MTs. Der Kontext zeigt jedoch eindeutig, dass es hier um explizit gelernte explizite MTs geht:

In der Fußnote zum nachfolgenden Satz, der den Extensionsunterschied der Klassen der „unlearned terms“ und der „innate terms“ behauptet, wird C. Danto (1969) mit Bezug auf das oben genannte Zitat zitiert, mit:

Für das explizite Lernen werden also explizite Hypothesen als Vorbedingung verlangt.

Der Lernprozess besteht dann bei Geltung der LOTH in einer empirischen Bestätigung („confirming“) dieser expliziten Hypothesen:

„Learning the theory will require formulating and confirming hypotheses such as these:

(A) Cats have whiskers.

(B) Cats have four legs.

(C) Cats have fur.

According to LOT, these hypothesis are represented as sentences in LOT.”<sup>45</sup> (Cummins 1997, S. 537).

Rupert behauptet nun mit seiner Theorie einen Weg aufgezeigt zu haben, wie solch ein expliziter Lernvorgang tatsächlich möglich wäre, ohne dass dieser dem Widerspruch zum Opfer fiele, den Cummins im zweiten Argumentationsschritt konstruiert hat.

In seiner Theorie des Lernens von MTs sieht Rupert eine wirkungsvolle Verteidigung der CT gegen Cummins’ Kritik im zweiten Argumentationsschritt.

Hierin ist jedoch eine Fehleinschätzung des Bezuges von Ruperts Theorie auf Cummins’ Argumentation zu sehen.

Bei Ruperts Theorie handelt es sich, wie ich im Folgenden zeigen werde, nicht um die Beschreibung eines unter Annahme der LOTH expliziten Lernvorgangs. Stattdessen, beschreibt Rupert einen bezüglich einer angenommenen LOT impliziten Lernvorgang.

Somit kann Rupert damit nicht, wie er behauptet („The alleged circularity of CT vanishes.“ (Rupert 2001, S. 524)), die von Cummins konstruierte Zirkularität bezüglich explizit gelernter MTs auflösen. Ruperts Theorie ist also kein Argument gegen Cummins’ zweiten Argumentationsschritt.

Stattdessen stellt Ruperts Theorie des Lernens ein weiteres Argument gegen das Ergebnis aus Cummins’ erstem Argumentationsschritt (Es gibt explizit gelernte MTs.) dar. Allerdings ist Ruperts Theorie des Lernens von MTs ein weitaus stärkeres Argument gegen Cummins’ Auffassung, es müsse explizit gelernte MTs geben und |cat|s MT gehöre dazu, als seine direkt gegen Cummins’ ersten Argumentationsschritt vorgebrachten Argumente. Denn hier zeigt Rupert tatsächlich einen Weg auf, auf welchem sogar explizit repräsentierte MTs implizit gelernt werden können.

Eine weitere Unklarheit in Ruperts Einschätzung der Wirkungsrichtung seiner Theorie des Lernens von MTs besteht im Hinblick auf die Frage, welche Versionen der CT aufgrund ihrer Rahmenannahmen von seiner Kritik ausgeschlossen sind.

Zu Recht verweist er darauf, dass seine Verteidigung der CT und LOTH unter den Grundannahmen, welche er „orthodox view“ nennt, nicht gültig wäre. Als der „orthodox view“ wird hier die Auffassung bezeichnet, dass für eine wissenschaftliche Beschreibung kognitiver Systeme gilt: Computationale Prozesse sind genau die Prozesse, die über semantisch interpretierten Symbolen definiert sind. Diese Auffassung habe ich in 2.3.2. (B) bereits erwähnt und diese Z. Pylyshyn zugeschrieben.

Zuzustimmen ist Rupert dahingehend, dass der von ihm so genannte „orthodox view“ von seiner CT Verteidigung ausgenommen ist. Aufgrund einer, wie ich gleich begründen werde, Fehlinterpretation einer Passage aus Pylyshyn (1989) ist Rupert jedoch der irrigen Auffas-

---

„Obviously, the move from *unlearned* to *innate* is abrupt, inasmuch as learning is but one avenue of acquisition ...’ [...].“ (Rupert 2001, S. 502, Fußnote 5). Somit ist mit „acquiring content“ tatsächlich der Lernprozess gemeint und „formulating an explicit theory“ ist also als Bedingung für den Lernprozess zu verstehen.

<sup>45</sup> Siehe Fußnote 15.

sung, seine Theorie des Lernens von MTs wäre mit Pylyshyns LOTH-Version, im Gegensatz zu Fodors, vereinbar.

Zur genaueren Bestimmung der Wirkungsrichtung und des Geltungsbereiches von Ruperts CT-Verteidigung werde ich also in diesem Abschnitt drei Thesen begründen:

- Ruperts Theorie des Lernens von MTs beschreibt einen impliziten Lernvorgang.
- Ruperts Theorie des Lernens von MTs stellt ein Argument gegen Cummins' ersten Argumentationsschritt dar.
- Ruperts Theorie des Lernens von MTs ist keine Verteidigung der LOTH Versionen von Fodor und Pylyshyn.

In den beiden Unterkapiteln 2.4.4. (A) und (B) werde ich dann untersuchen, ob Ruperts Theorie des Lernens von MTs inhaltlich schlüssig ist.

Dass es sich bei dem von Rupert beschriebenen Lernvorgang um implizites Lernen handelt, ist leicht zu begründen:

Ein expliziter Lernvorgang erfordert das Formulieren und das „confirming“ expliziter LOT-Sätze. Die Sätze „(A) *t* (a newly coined, nonsemantically individuated LOT term) have whiskers“ usw., welche Rupert als Kandidaten für einen Lernprozess vorschlägt, sind jedoch keine expliziten LOT-Sätze.

Die kompositionale Semantik ist einer der wesentlichsten Bestandteile der LOTH. Wenn die Semantik des Satzes (A) jedoch kompositional sein soll, muss der Gehalt des Satzes durch eine Funktion aus der Syntax des Satzes und den Gehalten aller Konstituenten bestimmt sein. „*t*“ hat jedoch in Ruperts Theorie ausdrücklich keinen Gehalt. Da einer der Konstituenten des Satzes (A) keinen Gehalt hat, kann gemäß der LOTH auch der Satz (A) selber keinen Gehalt haben. Es handelt sich somit bei den Sätzen der für das „confirming“ des Lernvorganges zur Verfügung stehenden Theorie nicht um explizite Repräsentationen in einem LOT-Schema.

Der von Rupert beschriebene Lernvorgang ist also kein expliziter Lernvorgang, sondern unter Annahme der LOTH allenfalls implizites Lernen.

In der Diskussion, ob die Sätze der Theorie außer „*t*“ noch weitere Protorepräsentationen<sup>46</sup> enthalten dürfen, schreibt Rupert einen impliziten Status nur solchen Theorien zu, welche völlig aus Protorepräsentationen bestehen. „But if other protorepresentations wholly mediate the context-fixing detection relations of a given protorepresentation *t*, the relevant axioms have the status only of an implicit theory of the things to which *t* will come to refer; [...]“ (Rupert 2001, S. 524, Fußnote 40).

Sätze, welche, wie die in seinem Beispiel benutzten, lediglich eine Protorepräsentation enthalten, bezeichnet Rupert hingegen als explizite Repräsentationen:

„[...] , for here I intend to show how a LOT term can be acquired by the mediating effects of *explicit* theory.“ (Rupert 2001, S. 524, Fußnote 40).

Damit ein komplexer LOT-Satz einen Gehalt bekommen kann, müssen aber aufgrund der kompositionalen Semantik von komplexen Repräsentationen nicht nur einige seiner Konstituenten, sondern alle Konstituenten einen Gehalt haben.

Eine Aneinanderreihung expliziter LOT-Grundterme und einer Protorepräsentation kann also aufgrund einer wesentlichen Grundaussage der LOTH prinzipiell nicht als expliziter LOT-Satz gelten.

Bei Geltung der LOTH ist Ruperts Theorie des Lernens von MTs eine Theorie impliziten Lernens und kann somit nicht als Kritik an Cummins' zweitem Argumentationsschritt gelten, da sich dieser nur gegen explizite Lernvorgänge richtet.

---

<sup>46</sup> Rupert bezeichnet die „nonsemantically individuated“ LOT-Terme, wie „*t*“ im oben genannten Beispiel auch als „protorepresentations“. Dieser Bezeichnung folge ich mit der Übersetzung „Protorepräsentation“.

Wie verhält sich nun Ruperts Theorie des Lernens von MTs tatsächlich zu Cummins' CT-Kritik?

Cummins' CT-Kritik im zweiten Argumentationsschritt wäre ohne das Ergebnis des ersten Argumentationsschrittes wirkungslos. Die Unmöglichkeit des expliziten Lernens von MTs bei Geltung der CT wäre kein Argument gegen die CT, wenn Cummins nicht schlüssig begründen kann, dass es bei Geltung der CT überhaupt MTs geben muss, die explizit gelernt sind.

Cummins' erster Argumentationsschritt ist bisher nur in seiner Methode, nicht in seinem Ergebnis direkt, angezweifelt worden.

Rupert hat versucht zu zeigen, dass bei Geltung der CT MTs auf anderem Wege entstehen können, als gelernt zu werden. Damit hat er den Bereich der nicht gelernten MTs, die Cummins noch auf im starken Sinne angeborene MTs beschränkt hatte, so erweitert, dass |cat|s MT möglicherweise dazu gehören könnte. Die Existenz von gelernten MTs wurde dabei so wenig in Frage gestellt, wie die Behauptung, dass gelernte MTs explizit gelernt werden müssten.

Ich habe in 2.3.2. (B) gezeigt, dass aus CT und LOTH keine Beschränkung der gelernten MTs auf explizite MTs folgt, dass CT und LOTH durchaus mit dem Lernen von impliziten MTs vereinbar sind.

Hierdurch habe ich das Ergebnis von Cummins' erstem Argumentationsschritt insofern angezweifelt, als dass ich gezeigt habe, dass man von der Notwendigkeit des Lernens bestimmter MTs nicht auf deren explizite Repräsentation schließen kann.

Die Argumente von Rupert und mir weisen also nur darauf hin, dass die Menge der impliziten MTs größer sein könnte, als Cummins zugesteht, dass diese nämlich weder angeboren im starken Sinne (Ruperts Argumente in 2.3.5.) noch ungelernt (meine Argumentation in 2.3.2. (B)) sein müssen.

Nicht behandelt wurde bisher die Frage, ob es überhaupt explizit repräsentierte MTs gibt, und wenn ja, ob diese explizit gelernt werden müssten.

Eine wirkungsvolle Verteidigung der CT durch Kritik an Cummins' erstem Argumentationsschritt müsste jedoch nicht nur für implizit repräsentierte MTs, sondern auch für explizit repräsentierte MTs zeigen, wie diese bei Geltung der CT und LOTH gelernt werden können, ohne dass hierfür explizites Lernen notwendig ist.

Genau in dieser Richtung wird die von mir in 2.3.2. (B) geübte Kritik an Cummins' erstem Argumentationsschritt durch Ruperts Theorie des Lernens von MTs ergänzt.

Ich habe dafür argumentiert, dass bei Geltung von CT und LOTH implizite MTs implizit gelernt werden können. Ruperts Theorie des Lernens von MTs zeigt, wie explizite MTs bei Geltung der CT und LOTH prinzipiell implizit gelernt werden können.

Sollte sich Ruperts Theorie des Lernens von MTs also in der in 2.4.4. (A) und (B) durchgeführten Untersuchung als schlüssig erweisen, wäre sie ein starkes Argument gegen Cummins' Ergebnis des ersten Argumentationsschrittes, aus LOTH und CT würde folgen, dass es MTs gäbe, die explizit gelernt werden müssten.

Welche CT-Versionen sind nun von Ruperts Verteidigung ausgenommen?

Rupert stellt sich selbst die kritische Frage: „[...] newly coined LOT terms would seem to be in the domain of some computational operations, despite the possibility that they represent nothing. Does not the orthodox computational approach to explaining cognition proscribe such models?“ (Rupert 2001, S. 526).

Der „orthodox view“ besteht in der Auffassung, dass kognitive Prozesse genau die Prozesse sind, welche über Repräsentationen zu beschreiben sind:

„[...] that it is *essential* to computation that it only be carried out over semantically interpreted states or units; [...]“ (Rupert 2001, S. 526, Fußnote 44).

Rupert beantwortet die Frage, ob der „orthodox view“ mit seiner Theorie des Lernens von MTs vereinbar ist, negativ: „[...] this understanding of computation seems to conflict with the present proposal.“ (Rupert 2001, S. 526, Fußnote 44).

Hierin ist Rupert Recht zu geben. Sofern Lernen als kognitiver Vorgang aufgefasst wird und Kognition durch Manipulation von Repräsentationen definiert wird, kann Ruperts Theorie des Lernens von MTs natürlich nicht als gültige Beschreibung eines Lernvorgangs bezeichnet werden.

Ich habe diese Grundannahmen über die Beziehung zwischen Kognition und Repräsentation bereits in 2.3.2. (B) im Zusammenhang mit Pylyshyns Theorie der Kognition besprochen.

Rupert schreibt diesen „orthodox view“ auch zunächst Fodor und Pylyshyn zu: „For statements of the orthodox view, see Fodor, *The Language of Thought*, pp. 31, 46, 56, 99; also see his *The Modularity of Mind* (Cambridge: MIT, 1981), p. 5, and *Concepts*, pp. 10-11, Pylyshyn, *Computation*, pp. 51-58; Churchland and Sejnowski, pp. 62-66.”<sup>47</sup> (Rupert 2001, S. 526 Fußnote 44).

Merkwürdigerweise meint Rupert jedoch, aus einer Äußerung Pylyshyns entnehmen zu können, dass Pylyshyn seinen „orthodox view“ soweit eingeschränkt hätte, dass dieser mit Ruperts Theorie des Lernens von MTs vereinbar wäre:

„[...] one might, as Pylyshyn does, think that our best cognitive theories will sometimes deal in computations over interpreted states, while at the same time including nonsemantic explanations of important aspects of cognition: ‘[W]e may need the biological level to explain other things as well, such as possibly the nature of cognitive development or maturation or psychopathology, and perhaps *some changes that are now called learning*; exactly what facts fall at each of the three levels [semantic, syntactic, and biological] remains to a large extent an open empirical question’ - ‘Computing in Cognitive Science,’ in Michael I. Posner, ed., *Foundations of Cognitive Science* (Cambridge: MIT, 1989), pp. 51-91, p. 61, italics added; [...].”<sup>48</sup> (Rupert 2001, S. 526, Fußnote 44).

Dieses Zitat von Pylyshyn ist von Rupert jedoch fehlinterpretiert worden. Pylyshyn will hier nicht, wie Rupert vermutet, andeuten, dass kognitive Prozesse, falls sich dies empirisch als vorteilhafter herausstellen sollte, gelegentlich auch als Prozesse über nicht semantisch interpretierten Symbolen beschrieben werden sollten.

Diese zitierte Aussage taucht bei Pylyshyn bereits in Pylyshyn (1984) in folgendem Zusammenhang auf:

In acht Kapiteln entwirft Pylyshyn dort seine Theorie der Kognition, welche unter anderem, um Eindeutigkeit bemüht, Kriterien aufstellt, für welche Gegenstände in einem kognitiven System die drei Beschreibungsebenen (biologisch, syntaktisch, semantisch) jeweils anzuwenden seien.

Einer der Kernpunkte dieser Theorie ist die Identifikation von Kognition mit Prozessen über Repräsentationen. Im letzten, mit „Epilogue: What Is Cognitive Science the Science of?“ überschriebenen, Kapitel versucht Pylyshyn kognitive Phänomene von solchen, in kognitiven Systemen zu beobachtenden, Phänomenen abzugrenzen, welche überhaupt keine kognitive Erklärung verlangen.

Das von Rupert genannte Zitat ist nicht als teilweise Zurücknahme von Pylyshyns Theorie zu verstehen, sondern als Aufzählung nichtkognitiver Prozesse, die zum Zustandekommen von kognitiv erklärbaren Prozessen beitragen. „the nature of cognitive development or maturation or psychopathology“ tauchen in Pylyshyns Theorie als nichtkognitive Veränderungen der funktionalen Architektur auf. Mit „some changes that are now called learning“ sind kognitive Prozesse gemeint, die noch (!) als Lernen bezeichnet werden. Diese sind hier im Sinne von

---

<sup>47</sup> Im Gegensatz zu Fodor (1975; 1983; 1998) und Pylyshyn (1984) wird der „orthodox view“ von Churchland et al. (1992) nur referiert aber nicht vertreten.

*Modularity of Mind* ist, abweichend von Ruperts Verzeichnis, 1983 erschienen.

<sup>48</sup> Eckige Klammern sind von Rupert gesetzt.

Kandidaten zu verstehen, welche sich als Ergebnis empirischer Forschung und richtiger Anwendung von Pylyshyns Theorie dann als nichtkognitive Phänomene herausstellen könnten und dann natürlich auch nicht mehr als „Lernen“ bezeichnet werden sollten.

Der Kontext in Pylyshyn (1989) legt ebenfalls diese Interpretation der betreffenden Passage nahe. Pylyshyns Theorie der Kognition beschränkt kognitive Phänomene eindeutig auf Prozesse über Repräsentationen. Der durch Rupert beschriebene Lernprozess wäre somit im Rahmen von Pylyshyns Theorie kein kognitiver Prozess. Da Pylyshyn den Begriff des Lernens auf ausschließlich kognitive Arten der systematischen, umweltbedingten Veränderung von Organismen beschränkt, könnte Ruperts Lernvorgang gemäß Pylyshyns Theorie nicht einmal als Lernvorgang bezeichnet werden. Eine den Kontext berücksichtigende Weise der Interpretation der zitierten Passage im Hinblick auf Ruperts Theorie könnte eher so aussehen: Falls ein Verhalten eines kognitiven Systems tatsächlich so beschrieben werden müsste, wie Rupert dies in seiner Theorie des Lernens von MTs tut, dann wäre dieses Verhalten nicht als kognitiver Prozess und somit auch nicht als Lernen zu bezeichnen; auch wenn dies noch (!) fälschlicherweise (!) geschieht („some changes that are now [!] called learning“).

Somit kann nicht angenommen werden, dass Pylyshyn von dem oben genannten „orthodox view“ soweit abweicht, dass er Ruperts Theorie des Lernens von MTs als Lernprozess akzeptieren würde. Also muss auch Pylyshyns Version der LOTH von Ruperts Verteidigung angenommen werden.

Insgesamt ist festzustellen, dass Ruperts Theorie des Lernens von MTs die Beschreibung eines impliziten Lernvorgangs für explizite MTs darstellt, und somit als zusätzliches Argument gegen Cummins' ersten Argumentationsschritt gelten kann. Dabei sind die CT- bzw. LOTH-Versionen, welche von Fodor und Pylyshyn vertreten werden, von Ruperts Verteidigung ausgeschlossen, da der von Rupert beschriebene Lernprozess mit den Rahmenannahmen dieser Versionen der LOTH nicht vereinbar ist.

Zwei Überlegungen sollen den Zusammenhang der drei eben begründeten Thesen verdeutlichen.

Erstens ist hier noch zu bemerken, dass Ruperts Theorie nur unter meiner Interpretation als Beschreibung eines impliziten Lernvorgangs als Verteidigung der CT und LOTH gelten kann: Mit folgender Bemerkung hat Rupert natürlich recht, da alle Teilnehmer der Diskussion ein empirisch erfolgreiches Modell anstreben: „[...] if our most empirically successful computational models of cognition include dummy terms, with syntactic and biological properties, but no semantical ones, then so be it.“ (Rupert 2001, S. 526, Fußnote 44).

Falls man Ruperts Theorie des Lernens von MTs jedoch wie er selbst als Beschreibung eines expliziten Lernvorgangs betrachtet, ist dies ein Argument gegen anstatt für die Geltung der LOTH. Ein Kernpunkt der allgemeinen Definition der LOTH ist es, dass eine komplexe Repräsentation – also auch ein expliziter Satz in einer zu überprüfenden Hypothese in Ruperts Lerntheorie – eben keine „dummy terms“ enthält, sondern Repräsentationen, deren Gehälter dann zusammen mit der Syntax der komplexen Repräsentation deren Gehalt bestimmen.

Wenn also ein kognitives Modell, welches Protorepräsentationen enthält, mit der LOTH vereinbar sein soll, dann dürfen diese Protorepräsentationen nicht als Bestandteile von expliziten Repräsentationen auftauchen.

Zweitens wirkt eine Theorie mentaler Repräsentationen, welche einen so wesentlichen Aspekt mentaler Repräsentationen wie das Lernen der Bildung von Repräsentationen ohne Rückgriff auf explizites Lernen erklären will, wohl auch auf liberalere Vertreter der LOTH als Fodor und Pylyshyn befremdlich. Wenn man die zu überprüfenden Hypothesen in Ruperts Theorie als implizite Repräsentationen auffasst, dann würde dies bedeuten, dass das kognitive System bei jedem Lernvorgang komplexe Strukturen bildet, welche dieselbe Syntax wie komplexe



Repräsentationen haben und zahlreiche echte Repräsentationen enthalten, aber selber keine komplexen Repräsentationen sind.

Die Annahme der LOTH wäre im Hinblick auf solch ein Modell sehr fragwürdig, da sie komplexen, syntaktisch völlig identischen Sätzen je nachdem, ob einer ihrer Konstituenten semantisch repräsentiert wäre oder nicht, insgesamt impliziten oder expliziten Charakter zuordnet.

Es ist also nicht zu vermuten, dass sich Ruperts Theorie des Lernens von MTs in der hier dargestellten Form in irgendeine ernsthaft vertretene LOTH bzw. CT-Version einbetten lässt.

Dass Ruperts Theorie als positive LOTH- bzw. CT-Theorie wenig plausibel klingt, beschädigt jedoch nicht deren argumentative Wirkung gegen Cummins' CT-Kritik.

Cummins möchte in Cummins (1997) schließlich nicht eine bestimmte Gruppe von CTs oder einen konkreten Vertreter kritisieren. Ihm geht es darum zu beweisen, dass die CT aus prinzipiellen Gründen, die aus ihrer allgemeinsten Definition abgeleitet werden können, falsch ist. „Since the argument uses only a generic form of CT, it follows that CT is false in all its forms.“ (Cummins 1997, S. 541).

Um diesen Ansatz anzugreifen genügt es einen allgemeinen Weg zu skizzieren, auf dem eine CT Cummins' Argumenten entgegen gehen kann. Sollte sich Ruperts Theorie des Lernens von MTs also im Folgenden als inhaltlich konsistent erweisen, kann diese als Argument gegen Cummins' CT Kritik gelten, ohne dass Rupert damit eine ausgereifte, plausible Version der CT beschrieben hätte.

#### 2.4.4. (A) Funktionsweise des „coining“

Das Problem, wie LOT-Symbole physisch entstehen können, stellt sich nicht nur für Ruperts Theorie des Lernens von MTs, sondern grundsätzlich für alle Versionen der LOTH. Des Weiteren kritisiert Cummins die CT/ LOTH hier nicht durch Überlegungen bezüglich der physischen Realisierung oder Entsehung von LOT-Symbolen, sondern ausschließlich durch Argumente zu MTs von LOT-Termen. So ist es für eine Verteidigung der CT gegen Cummins' Argument auch irrelevant, ob Rupert dieses Problem zufriedenstellend löst. Ich habe deshalb Ruperts Lösungsvorschläge dazu auch schon in 2.4.3. (A) nicht inhaltlich dargestellt und werde diese auch hier nicht diskutieren.

Stattdessen werde ich mich jetzt der interessanteren Frage zuwenden, wie das „coining“ ausgelöst werden soll, und dabei voraussetzen, dass es prinzipiell möglich ist, dass LOT-Symbole auf physikalisch beschreibbare Weise entstehen können.

Merkwürdig erscheint hier zunächst, dass Rupert zur Verteidigung der CT/ LOTH eine Theorie entwirft, welche „cluster“ von Eigenschaften und „Ähnlichkeitsräume“ enthält.

Diese Begriffe erinnern eher an ein repräsentationales Schema ähnlich der „state space semantics“, welches von z.B. P. M. Churchland vertreten wird. Diese Theorie ist von LOTH-Theoretikern scharf kritisiert worden; z.B. in einer Debatte zwischen Churchland (1989, 1996a, 1996b) und Fodor; Lepore (1992, 1996), wobei die Unvereinbarkeit von LOTH und „state space semantics“ auf beiden Seiten angenommen wird.

Die Zitate in 2.4.3. (A) zeigen jedoch eindeutig, dass Rupert der Auffassung ist, man könne die MT eines LOT-Grundterms sowohl als Menge von LOT-Sätzen als auch als „cluster“ von Eigenschaften beschreiben.

Setzt man dies voraus, folgen daraus einige Einschränkungen für die zu lernende MT eines LOT-Grundterms. Diese Bedingungen werde ich anhand folgender Fragen herleiten und bezüglich Ruperts gesamter Theorie des Lernens von MTs bewerten:

- Was muss für einen „Ähnlichkeitsraum“ gelten, damit er eine LOT-MT ausdrücken kann?
- Wie sieht dann eine LOT-MT in solch einem „Ähnlichkeitsraum“ aus?

- Wie ist unter diesen Bedingungen das „coining“ zu interpretieren?
- Kann eine solche Theorie als Verteidigung der CT gegen Cummins' Kritik gelten?

Was muss für einen „Ähnlichkeitsraum“ gelten, damit dieser eine LOT-MT ausdrücken kann? In der „state space semantics“ wird ein Ähnlichkeitsraum als metrischer Raum betrachtet, dessen Dimensionen Eigenschaften repräsentieren. Die Werte in einer Dimension repräsentieren quantitative Ausprägungen der Eigenschaft. Man kann nun Objekte oder Prototypen als Tupel von quantitativ beschriebenen Ausprägungen ihrer Eigenschaften durch Punkte in so einem Raum beschreiben. Für die Beschreibung von „Begriffen“ kann man auch Intervalle oder Intervalle mit Wahrscheinlichkeitsfunktionen definieren. Eine geeignete Metrik des Raumes wird als „Ähnlichkeit“ interpretiert.

Wenn nun eine Menge von LOT-Sätzen,<sup>49</sup> wie sie von Rupert beschrieben wird, durch einen solchen Raum dargestellt wird, ist es nahe liegend, die Eigenschaften, deren Ausprägungen in den Dimensionen ausgedrückt werden sollen, mit den Eigenschaften zu identifizieren, welche der „Protorepräsentation“  $t$  von den Sätzen zugeschrieben werden. Heißt ein Satz in der MT „(1)  $t$  have whiskers“, dann entspricht die Eigenschaft, Schnurrhaare zu haben, einer Dimension in dem entsprechenden Raum.

Welche Werte kann nun eine MT bezüglich dieser Dimensionen haben?

Eine positive Eigenschaft von solchen Ähnlichkeitsräumen ist, dass man für Objekte und Begriffe verschiedene Ausprägungen von Eigenschaften ausdrücken kann.

Für Räume, welche LOT-MTs darstellen sollen, kommt dies jedoch nicht in Frage. Die von Rupert und Cummins erwähnten Sätze, wie „ $t$  have whiskers“, behaupten, dass einer Protorepräsentation eine Eigenschaft zukommt, nicht dass ihr eine Eigenschaft zu einem gewissen Grad zukommt. Also können die Dimensionen lediglich zwei verschiedene Werte annehmen. Einen, der ausdrückt, dass die Protorepräsentation die Eigenschaft hat, und einen, welcher ausdrückt, dass die Protorepräsentation die Eigenschaft nicht hat.<sup>50</sup>

Die Bezeichnung der wahrzunehmenden Objekte und MTs als „cluster“ suggeriert des Weiteren eine gewisse Unschärfe. Tatsächlich wird es auch allgemein als eine vorteilhafte Eigenschaft von Ähnlichkeitsräumen angesehen, die Unschärfe von Begriffen, durch Wahrscheinlichkeitsfunktionen auf diesen Räumen auszudrücken.

Auch dies kommt für LOT-Sätze, wie bei Rupert beschrieben, nicht in Frage: „ $t$  have whiskers“ bedeutet eben nicht, „ $t$  hat mit der Wahrscheinlichkeit  $p=0,6$  Schnurrhaare.“

Der Ähnlichkeitsraum für die Darstellung von expliziten LOT-MTs hat also so viele Dimensionen, wie Eigenschaften detektiert werden können, und nimmt pro Dimension einen von zwei Werten an.

Die Metrik des Raumes ist dadurch natürlich sehr eingeschränkt. Eine sinnvolle Messung der Ähnlichkeit kann in diesem Rahmen eigentlich nur noch in einer Abzählung der von MT und Objektwahrnehmung übereinstimmend zugeordneten Eigenschaften bestehen.

---

<sup>49</sup> Ich bezeichne hier die von Rupert definierten Ausdrücke, wie „ $t$  have whiskers“ als LOT-Sätze, obwohl es, wie ich oben gezeigt habe, eigentlich implizite Ausdrücke sind. Alle bezüglich der Ähnlichkeitsräume zu nennenden Probleme treffen aber wegen deren Ähnlichkeit zu expliziten LOT-Sätzen auch auf die von Rupert benutzten impliziten Sätze zu.

<sup>50</sup> Dass ein Objekt oder eine MT eine Eigenschaft nicht instanziiert, fügt zu einer LOT-Satz Menge, wie von Rupert beschrieben, zwar Informationen hinzu. MTs und Objekte sollen jedoch innerhalb eines Raumes verglichen werden. Somit wäre die Benutzung des Begriffs „Ähnlichkeitsraum“ völlig sinnlos, wenn jede MT einen eigenen Raum, mit Anzahl der Dimensionen gleich Anzahl der Sätze, definieren würde. Somit muss hier angenommen werden, dass Objekte und MTs als „cluster“ in einem einheitlichen Ähnlichkeitsraum aufgefasst werden, dessen Dimensionen alle Eigenschaften sind, welche das System im Moment des Vergleichs überhaupt detektieren kann.

Da also offenbar durch die Beschreibung als „cluster“ von Eigenschaften der Beschreibung als LOT-Satzmenge keinerlei Ausdrucksmöglichkeiten hinzugefügt werden, fragt sich, warum Rupert überhaupt auf diesen begrifflichen Rahmen zurückgreift.

Vermutlich weil der nichtsemantische Charakter des Vergleichs zwischen MTs mit dem Ausdruck „Ähnlichkeitsmessung“ deutlicher erscheint als bei einer Beschreibung im LOTH-Vokabular.

Wie sieht eine LOT-MT in so einem Ähnlichkeitsraum aus?

Für die MT als LOT-Satzmenge bringt es ebenso Einschränkungen mit sich, in einem Ähnlichkeitsraum beschreibbar zu sein.

Erstens kann die Theorie tatsächlich nur Sätze enthalten, welche einer Protorepräsentation auf die Weise Eigenschaften zuordnen, wie die von Rupert als Beispiel genannten Sätze.

Ein wesentlicher Einwand der Kritiker der „state space semantics“ ist, dass man in einem Eigenschaftsraum keine Kompositionalität ausdrücken kann. „[...], concepts *can't* be prototypes – [...] – because prototypes aren't compositional and concepts are.“ (Fodor; Lepore 1996, S. 162).

In 1.2.3. und 2.3.2. habe ich bereits Begründungen aus Cummins (1996b) und Cummins et al. (2001) dafür genannt, dass es tatsächlich nicht möglich ist, durch allgemeine metrische Funktionen in einem Ähnlichkeitsraum LOT-Kompositionalität auszudrücken.

Deshalb darf die MT keine komplexen Sätze enthalten (Strenggenommen werden im Ähnlichkeitsraum auch nicht die grammatischen Eigenschaften von „*t* have whiskers“ dargestellt, sondern lediglich die zugeordnete Eigenschaft.).

Zweitens dürfen die MTs aus demselben Grund auch keine komplexen Eigenschaften zuordnen, sondern nur LOT-Grundterme enthalten.

Zum Preis der Ausdrückbarkeit als Punkt in einem rudimentären Ähnlichkeitsraum muss also bei den möglichen Sätzen der MT auf die Haupteigenschaft von LOT-Sätzen – die Kompositionalität – verzichtet werden.

Da die wichtigsten Ausdrucksmöglichkeiten von „state space semantics“ und LOTH – quantitative und statistische Darstellung von Begriffen und Kompositionalität – nicht ohne weiteres vereinbar sind, muss auf diese verzichtet werden, falls eine MT sowohl in einem Eigenschaftsraum als auch in einem LOT-Schema zu beschreiben sein soll.

Wie ist unter diesen Bedingungen das „coining“ zu interpretieren?

Sind in einem wahrzunehmenden Objekt Eigenschaften instanziiert, welchen im kognitiven System ein LOT-Grundterm entspricht, dann werden nach CT und LOTH die entsprechenden LOT-Grundterme gebildet („tokening“)<sup>51</sup>. Nach Ruperts Theorie werden nun die Menge<sup>52</sup> der gebildeten LOT-Grundterme und/oder ihre Teilmengen mit allen expliziten MTs auf Ähnlichkeit hin verglichen. Fällt der Vergleich für die Menge der gebildeten LOT-Grundterme (oder für eine Teilmenge) bei allen MTs negativ aus, dann wird ein neues LOT-Symbol gebildet und mit einer expliziten MT ausgestattet, welche dem LOT-Symbol die gebildeten Repräsentationen zuordnet.

Ob auch die Teilmengen verglichen werden sollen oder nicht, bleibt bei Rupert offen. Beides wäre erklärungsbedürftig. Würde nur die Gesamtmenge verglichen, würde dies zu erheblichen Redundanzen führen (welche durch den Lernprozess nicht beseitigt würden. Siehe 2.4.4. (B)):

---

<sup>51</sup> Nach Kapitel 2.2.4. wird von Cummins und Rupert übereinstimmend nicht vorausgesetzt, dass komplexe Eigenschaften direkt detektiert werden.

<sup>52</sup> Aufgrund der oben beschriebenen minimalen Bedeutung des Begriffs Eigenschafts-„cluster“ bezeichne ich die gebildeten MTs einfach als Menge.

Enthält eine MT für eine Protorepräsentation  $a$  den Satz „ $a$  hat die Eigenschaft, eine Katze zu sein“ enthält sie auch die Sätze „ $a$  hat Schnurrhaare.“, „ $a$  hat vier Pfoten.“, usw. .

Würde die Menge der Teilmengen zum Vergleich herangezogen, würde dies dazu führen, dass zu einem LOT-Term mit expliziter MT bei jedem „tokening“ zahlreiche, vielleicht unnütze LOT-Symbole gebildet würden, deren zu lernende MTs dann aus Teilmengen der LOT-Sätze bestünden, welche die explizite MT des Terms konstituieren.

Ruperts Beschreibung des „coining“ verlangt also für beide Möglichkeiten genauere Erklärungen.

Kann eine solche Theorie als Verteidigung der CT gegen Cummins' Kritik gelten?

Erstens wirkt Ruperts Theorie recht unausgereift. Die Beschreibung der MTs als Eigenschaftsräume, sowie der Vorgang des „coining“ selbst, werden nur in Andeutungen genauer erklärt, werfen jedoch schon auf den ersten Blick einige Fragen auf.

Zweitens werden von Rupert hier Theorieelemente bemüht, welche eher in Theorien zu finden sind, die zur LOTH im klaren Widerspruch stehen („culster“, Ähnlichkeitsräume, ...).

Drittens wird in der Formulierung der Theorie nirgends auf die LOTH Bezug genommen.

Deshalb erscheint es eher unwahrscheinlich, dass ein LOTH-Theoretiker diese Theorie in dieser Form positiv vertreten würde.

Allerdings steht die Theorie auch nicht im Widerspruch zur CT und LOTH. So werden die „state space semantics“ – Elemente lediglich für die Beschreibung von Mengen von Eigenschaften benutzt und nicht für komplexe Sätze.

Somit kann das „coining“, falls es sich in eine Lerntheorie einbetten lässt, durchaus in einem Argument gegen Cummins verwendet werden.

#### 2.4.4. (B) Lernen mit „coined terms“

Nach einer Bemerkung zu Ruperts Lernprozess im allgemeinen werde ich in diesem Kapitel die Verträglichkeit von Ruperts Theorie mit der LOTH besprechen. Dann werde ich die in 2.4.3. (A) angesprochene Problematik der Gehaltszuschreibung weiter diskutieren und abschließend noch einmal auf die Rolle der Theorie als Verteidigung gegen Cummins' Argument eingehen.

Der Lernprozess allgemein:

Zwar beschreibt Rupert die Lernschritte für seine Lerntheorie als „Hinzufügungen und Ändern von Axiomen“, jedoch sollen diese Änderungen auch als „pulling in similarity space“ aufzufassen sein. Da für die Lernschritte dieselbe Regel/Theorie wie für das „coining“ zuständig ist, ergeben sich dieselben Probleme, die ich in 2.4.4. (A) schon für das „coining“ angesprochen habe:

- Die LOT-MTs sind in ihrer Struktur, aufgrund der Beschreibbarkeit als Punkte im Ähnlichkeitsraum, sehr eingeschränkt.
- Genau wie die Einführung der MTs lassen auch die Veränderungen im Lernschritt einige problematische Fragen offen.

Ruperts Lerntheorie und die LOTH:

Ruperts Theorie verzichtet auf eine Erklärungsebene, die von LOTH-Theoretikern oft als zentrale Leistung ihrer Theorie gesehen wird.

Als eine wichtige Eigenschaft der LOTH wird von deren Vertretern oft bezeichnet, dass die LOTH ermöglicht, die Interaktionen eines kognitiven Systems mit seiner Umwelt als rationale Prozesse beschreiben zu können.

Bezüglich des Lernens von Repräsentationen entspricht dies auch der Intuition, dass das Generieren und Verbessern von Repräsentationen zumindest in einigen Fällen als rationaler Vorgang empfunden wird, und nicht als quantitativ festgelegte Funktion von Mengen von Eigenschaftsinstanzierungen in der Umwelt.

Insbesondere das Lernen von Wörtern bei Erwachsenen, wozu schließlich auch das Lernen von Repräsentationen der Wörter gehört, legt diese Intuition nahe.

Diese von LOTH-Theoretikern oft in Anspruch genommene explanatorische Leistung kann von Rupert nicht erfüllt werden.

Da Cummins in seinem Argument die CT nicht wegen einer Unfähigkeit Lernprozesse rational zu beschreiben angereift, beschädigt diese Schwäche hier zwar nicht die argumentative Wirkung von Ruperts Theorie gegen Cummins' CT-Kritik, jedoch unterstützt dies den bereits in 2.4.4. (A) geäußerten Eindruck, dass seine Theorie wohl kaum in dieser Form als positive CT-Theorie vertretbar wäre.

Zuschreibung des Gehaltes:

Ein schwerwiegender Einwand gegen Ruperts Theorie ist, dass er zwar eine Erklärung dafür skizziert, wie Protorepräsentationen entstehen und ihre MTs verändern können, aber nicht, wie sie Gehalt bekommen können.

Eine Erklärung hierfür wird bei Rupert nicht einmal im Ansatz angedeutet. Es erscheint im Rahmen seiner Lerntheorie sogar fraglich, warum es überhaupt Repräsentationen geben sollte, da kein Grund dafür angedeutet wird, warum die Protorepräsentationen, wenn ihre MT zu einem zufriedenstellenden Grad verbessert ist, überhaupt zu Repräsentationen umgewandelt werden sollten.

Cummins deutet ein ähnliches Problem bereits in seinen Repliken auf potentielle Einwände für die Lernschritte an:

Cummins schließt dort Theorien mit Protorepräsentationen aus, da sie, wie er meint, |cat| brauchen um den Unterschied zwischen |cat| und *t* zu repräsentieren. „Why should the theory become elaborated? There is not, after all, an error signal that says, in effect, that the theory is not right yet. [...] to elaborate the theory in the right direction, you have to know that the theory is not yet a good enough theory of catness, and that, it would seem, requires having |cat|s in your repertoire, [...]” (Cummins 1997, S. 538).

Cummins kritisiert, wie in vielen seiner Kommentare zu CT und LOTH (siehe 1.2.2.), die CT mit der Behauptung, sie könne keine repräsentationalen Fehler und damit kein Lernen erklären. So reagiert Rupert auch folgendermaßen: „Note that, if a new experience causes a change in theoretical axioms in which *t* appears, this does not imply that the cognitive system explicitly represented *error* in its earlier theory of *t*.” (Rupert 2001, S. 525).

Die Frage, ob die CT Fehler zulässt, muss unabhängig von der hier dargestellten Debatte behandelt werden, da dies Cummins' Argument eine völlig andere Struktur geben würde (siehe Fußnote 18). Somit werde ich hier nicht diskutieren, ob Ruperts Rechtfertigung gegen diesen Einwand berechtigt ist.

Was Cummins hier gegen die Möglichkeit vorbringt einen Lernprozess zu beginnen, gilt jedoch auch für die Möglichkeit den Lernprozess zu beenden.

Zwar muss nicht unbedingt gefordert werden, dass das System schon |cat| besitzt, um den Lernprozess zu beenden, aber eine Bedingung für die letztendliche Zuschreibung der Gehaltes muss schon formuliert werden.

Wie das System „weiß“, dass die Theorie nicht gut genug ist, wird von Rupert zumindest beantwortet. Man könnte aber genauso gut fragen, woher das repräsentationale System „weiß“, dass die MT bereits gut genug ist, und die Protorepräsentation zur Repräsentation werden kann.

Da die „coinig“/ „pulling“- Regel nur quantitative Eigenschaften der MT einer Protorepräsentation verarbeiten kann, müsste die Beschreibung einer solchen Umwandlung Ruperts Theorie ein substantiell neues Element hinzufügen.

Rupert nimmt für seine Theorie in Anspruch zu zeigen, wie einer Repräsentation Gehalt zugewiesen wird.

„Even where explicit theory contributes to the fixation of content, there exists a possible route to content fixation within the framework assumed by CT, a route Cummins fails to consider.“ (Rupert 2001, S. 530).

Tatsächlich gibt er genau für diese Frage – wie erhält eine Repräsentation Gehalt? – nicht einmal den Ansatz einer Antwort, sondern weist diese einer genaueren Formulierung der CT zu:

„This suggests a condition that must be met by a causal theory of content: it must assign content to newly coined, content-lacking LOT terms in a timely enough fashion that such terms possess content when our explanatory purposes demand that they do.“ (Rupert 2001, S. 527, Fußnote 47).

Bei dieser Bedingung handelt es sich nicht, wie Rupert vielleicht suggerieren möchte, um ein technisches Randproblem, sondern um die absolut zentrale Frage, die eine Theorie des Lernens repräsentationaler MTs beantworten muss.

Somit kann Ruperts Theorie des Lernens von MTs nicht als eine Theorie aufgefasst werden, welche das Lernen der MT einer Repräsentation erklärt, sondern lediglich als eine rudimentäre Skizze der Entstehung und Veränderung von nicht repräsentierenden Symbolen.

Ruperts Theorie und Cummins' Argument:

Zu Anfang des Kapitels 2.4.4. habe ich dargelegt, dass Ruperts Theorie des Lernens von MTs entgegen seiner eigenen Darstellung nicht als Kritik an Cummins' zweitem Argumentationsschritt aufzufassen ist, sondern als klare Kritik an Cummins' erstem Argumentationsschritt. Ruperts Theorie des Lernens von MTs muss als Beschreibung eines impliziten Lernvorgangs von expliziten MTs aufgefasst werden und richtet sich somit gegen Cummins' Implikationsbehauptung – gelernte MTs sind explizit gelernte und explizit repräsentierte MTs – und damit gegen seinen ersten Argumentationsschritt.

Da sich Ruperts Theorie durch die Untersuchung in diesem Kapitel jedoch als unzulänglich für die Erklärung des Lernens von expliziten MTs von Repräsentationen erwiesen hat, wird von seiner Theorie weder Cummins' erster noch Cummins' zweiter Argumentationsschritt wirksam beschädigt.

### **III. Fazit**

In diesem Kapitel werde ich zuerst das Ergebnis der in II. durchgeführten Untersuchung in einer abschließenden Beurteilung von Cummins' Argument gegen die CT zusammenfassen. Nach einer zusammenfassenden Bewertung von Ruperts CT-Verteidigung und deren Bezug zu Cummins' Argument werde ich dann abschließend eine durch die in II. durchgeführte Untersuchung motivierte Überlegung zu Cummins' CT-Kritik im Allgemeinen anstellen.

Zu Cummins' Argument:

Cummins' hier besprochenes Argument gegen die CT ist von mir nach zwei Argumentationsschritten getrennt untersucht worden. Dabei hat sich herausgestellt, dass bezüglich Cummins' erstem Argumentationsschritt sowohl die Prämissen, als auch das Ergebnis seines Schlusses zu bezweifeln sind.

In 2.3.3. bis 2.3.6. habe ich dargelegt, wie Rupert zurecht darauf hinweist, dass Cummins' Prämissen nicht begründet sind.

Eine Prämisse – die Implikationsbehauptung, Lernen impliziert, bei Geltung von CT und LOTH, für MTs explizite Repräsentation und explizites Lernen – habe ich in 2.3.2. widerlegt, indem ich gezeigt habe, dass MTs durchaus implizit gelernt werden können, ohne die Minimaldefinitionen von CT und LOTH zu verletzen.

Dadurch ist das Ergebnis von Cummins' erstem Argumentationsschritt – es gibt explizit gelernte MTs – bezweifelt worden. Warum sollte für die Entstehung von MTs das problematische explizite Lernen angenommen werden, wenn man das Lernen von MTs prinzipiell auch implizit erklären kann?

Der zweite Argumentationsschritt hat sich in der Untersuchung als richtig erwiesen. Eine der Prämissen – die Implikationsbehauptung, Lernen impliziert, bei Geltung von CT und LOTH, für MTs explizite Repräsentation und explizites Lernen – ist zwar, wie ich schon in der Kritik am ersten Argumentationsschritt gezeigt habe, allgemein zu bezweifeln. Bezüglich der MTs, auf welche sich der zweite Argumentationsschritt bezieht, ist die Prämisse jedoch trivialerweise wahr, da es sich schließlich gerade um explizit gelernte MTs handelt, deren Existenz im ersten Argumentationsschritt gezeigt werden sollte.

Von Ruperts Kritik wird Cummins' zweiter Argumentationsschritt schon deshalb nicht getroffen, weil diese Kritik, wie ich in 2.4.4. gezeigt habe, im Erfolgsfall lediglich das Ergebnis des ersten Argumentationsschrittes angreifen würde.

Da das Ergebnis von Cummins' erstem Argumentationsschritt zu bezweifeln ist und der im zweiten Argumentationsschritt konstruierte Widerspruch ohne dieses Ergebnis ins Leere läuft, kann Cummins' Argument insgesamt nicht als gültiges Argument gegen die CT angesehen werden.

Es ist der CT in ihrer Minimalversion also nicht aus den von Cummins in Cummins (1997) vorgebrachten Gründen prinzipiell unmöglich, die Entstehung von Repräsentationen durch Lernen zu erklären, selbst wenn man annimmt, dass die CT die LOTH impliziert.

Es sei hier noch einmal darauf hingewiesen, dass die Version der LOTH und CT, wie sie von Pylyshyn vertreten wird, tatsächlich Cummins' Argument zum Opfer fällt. In 2.3.2. habe ich darauf hingewiesen, dass meine Verteidigung der allgemeinen Version der LOTH und CT bei Annahme von Pylyshyns Beschreibungsmodell für kognitive Systeme (siehe 2.2.1., 2.3.2. (B) und 2.4.4.) ungültig wäre.

Somit kann Cummins' Argument durchaus als schlüssiges Argument gegen Pylyshyns LOTH und CT-Version angesehen werden.

Die allgemeine Version der CT wird dadurch jedoch nicht widerlegt.

Zu Ruperts CT-Verteidigung:

Ruperts direkte Kritik an Cummins' erstem Argumentationsschritt hat sich durch die Untersuchung in 2.3.6. als wenig wirkungsvoll herausgestellt. Cummins' Ergebnis wird von Rupert nicht in Zweifel gezogen. Statt dessen zeigt Rupert mit erheblichem argumentativen Aufwand – der Teil nimmt in Rupert (2001) 20 Seiten ein (S. 501 bis S. 521) – was anhand des Umfangs der Darstellung bei Cummins – zwei Sätze und Fußnoten (S. 536 f.) – ohnehin zu vermuten war: Cummins hat die Voraussetzungen seines ersten Argumentationsschrittes nicht plausibel begründet.

Cummins' Ergebnis wird jedoch von Rupert ausdrücklich akzeptiert (siehe 2.3.6. (F) und 2.4.3.).

Als Gegenargument zu Cummins' zweitem Argumentationsschritt skizziert Rupert eine eigene Theorie, welche erklären soll, wie MTs bei Geltung der CT und LOTH gelernt werden können.

Erstens beschreibt diese Theorie, wie ich in 2.4.4. gezeigt habe, einen impliziten Lernvorgang und ist somit eigentlich als Kritik an Cummins' erstem Argumentationsschritt aufzufassen. Zweitens kann Ruperts Theorie das Lernen von MTs von Repräsentationen, wie ich in 2.4.4. (B) gezeigt habe, unabhängig von der Geltung der CT und LOTH, nicht erklären.

Somit ist weder Ruperts direkte Kritik an Cummins' erstem Argumentationsschritt noch seine eigene Theorie des Lernens von MTs als wirkungsvolle Verteidigung der CT gegen Cummins' Kritik anzusehen.

Lernen bei Cummins:

Durch die Untersuchung von Cummins' Argument gegen die CT in Cummins (1997) und Ruperts Verteidigung der CT in Rupert (2001) ist ein generelles Problem bezüglich Cummins' Ansätzen zur CT-Kritik besonders deutlich zutage getreten.

Cummins kritisiert die CT oft – z.B. in Cummins (1996a; 2002a; forthcoming) – über Argumente zur Möglichkeit des repräsentationalen Lernens.

Jedoch gibt Cummins nirgends eine allgemeine Charakterisierung eines Lernprozesses oder stellt eine eigene Theorie repräsentationalen Lernens auf.

Als einzige positive Bedingung für eine Theorie repräsentationalen Lernens gibt Cummins meist die Darstellbarkeit eines repräsentationalen Fehlers an, um dann dafür zu argumentieren, dass die CT diese Bedingung nicht erfüllen kann. Wie ein solcher Fehler dann allerdings zum Lernen von Repräsentationen beitragen soll, bleibt bei Cummins stets offen.

Es wäre für eine Untersuchung der Lernbarkeit von MTs unter Voraussetzung der CT und LOTH allerdings zweckdienlicher gewesen, Forderungen für einen möglichst allgemeinen Lernbegriff zu formulieren und dann zu überprüfen, ob diese Forderungen mit der allgemeinen Form der CT zu vereinbaren sind.

Stattdessen formuliert Cummins in Anlehnung an Pylyshyns Beschreibungsmodell eine sehr strenge Definition von repräsentationalem Lernen und setzt diese (fälschlicherweise, wie ich in 2.3.2. gezeigt habe) dann als für die CT repräsentativ und unumgänglich voraus.

In Ruperts Verteidigung der CT herrscht eine ähnliche Unklarheit bezüglich dessen, was eigentlich erklärt werden soll, vor.

Es wäre hier sinnvoll gewesen, zuerst Klarheit darüber zu schaffen, was überhaupt, unabhängig von CT und LOTH, als Lernprozess gelten kann, und dann zu zeigen, wie ein solcher Prozess im Einklang mit CT und LOTH beschrieben werden könnte.

Stattdessen akzeptiert Rupert zuerst Cummins' unsinnige Einschränkung von Lernen auf explizites Lernen um dann eine Theorie zu entwerfen, welche dieser Definition nicht gerecht wird.

Ich habe in der vorliegenden Arbeit lediglich belegt, dass die von Cummins in Cummins (1997) gegen die CT vorgebrachten Argumente, obwohl sie durch Ruperts Kritik nicht getroffen werden, nicht (durch Nachweis der Unlernbarkeit von MTs) die CT widerlegen können.

Die Frage, ob die allgemeine CT nicht vielleicht noch andere, hier nicht berührte, schwerwiegende Probleme bezüglich des Lernens von MTs aufweist, ist hier nicht behandelt worden. Somit ist das Problem der Lernbarkeit von LOT-Termen hier durch die Analyse der beiden Texte von Cummins und Rupert noch nicht erschöpfend untersucht worden.

Für eine weiterführende Untersuchung der Lernbarkeit von Repräsentationen bei Geltung der CT wäre es jedenfalls äußerst sinnvoll, zuerst möglichst allgemeine Bedingungen dafür anzugeben, was ein Lernprozess sein soll.



Cummins' Forderung, dass in einem Lernprozess irgendwie ein repräsentationaler Fehler vorkommt, ist für solch eine Charakterisierung sicherlich zu wenig: Dass Fehler nicht mit einem Lernprozess gleichzusetzen sind, ist ein bekanntes Problem der Pädagogik.

Die Bedingungen, welche Cummins, motiviert durch Pylyshyn, als Folgerungen aus der LOTH und CT betrachtet – explizites Lernen – sind sicherlich, sogar für die LOTH, viel zu stark. Dies habe ich in 2.3.2. (B) gezeigt.

Es bedarf also einer Definition, welche die Vorgänge, die in den Kognitionswissenschaften sinnvoll als Lernen bezeichnet werden, gegen unproduktive Veränderungen und „irgendwie Fehler enthaltende“ Prozesse abgrenzt aber noch keine Implikationen einer bestimmten Theorie mentaler Repräsentationen voraussetzt.

Meine Untersuchung zeigt jedenfalls deutlich (besonders in 2.3.2., 2.3.6. und 2.4.4.), dass die Hauptprobleme bezüglich der Diskussionsführung – Ruperts konfuse Argumentationsstruktur in 2.3.5 und Cummins' unreflektierte Einführung des expliziten Lernens – durch das Fehlen einer solchen Definition bedingt sind.

Hieraus folgt, dass einer weiterführenden Diskussion der Lernbarkeit von Repräsentationen auf jeden Fall eine geeignete allgemeine Definition eines Lernprozesses vorausgehen sollte.

## Literaturverzeichnis:

- Beckerman, A. (2001): *Analytische Einführung in die Philosophie des Geistes*. Berlin; New York: de Gruyter; Zweite Auflage.
- Carrier, M. (1996): *Sprache des Denkens*. in Mittelstraß, J. (Hrsg.): *Enzyklopädie Philosophie und Wissenschaftstheorie*. Stuttgart; Weimar: Metzler.
- Changeaux, J. P. (1997) : *Neuronal Man*. Princeton : University Press; Zweite Auflage.
- Churchland, P. M. (1989): *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge, Mass.: MIT Press.
- Churchland, P. M. (1996a): *Fodor and Lepore: State-Space Semantics and Meaning Holism*. in McCauley, R. N. (Hrsg.): *The Churchlands and their Critics*. Cambridge, Mass.: Blackwell.
- Churchland, P. M. (1996b): *Second reply to Fodor and Lepore*. in McCauley, R. N. (Hrsg.): *The Churchlands and their Critics*. Cambridge, Mass.: Blackwell.
- Churchland, P. S.; Sejnowsky, T. J. (1992): *The Computational Brain*. Cambridge, Mass.: MIT Press.
- Clapin, H. (2002): *Tacit Representation in Functional Architecture*. in Clapin, H.: *Philosophy of Mental Representation*. Oxford: Clarendon Press.
- Cummins, R. (1986): *Inexplicit Information*. in Brand, M.; Harnish, R. M.(Hrsg.) : *The Representation of Knowledge and Belief*. Arizona University Press.
- Cummins, R. (1996a): *Representations, Targets, and Attitudes*. Cambridge, Mass.: MIT Press.
- Cummins, R. (1996b): *Systematicity*. in *Journal of Philosophy* Vol. 93.
- Cummins, R. (1997): *The LOT of the Causal Theory of Mental Content* in *Journal of Philosophy* Vol. 94.
- Cummins, R. (2000): *Reply to Milikan*. in *Philosophy and Phenomenological Research* Vol. 60.
- Cummins, R. (2002a): *Truth and Meaning*. in Keim-Campbell, J.; O'Rourke, M.; Shier, D. (Hrsg.): *Meaning and Truth: Investigations in Philosophical Semantics*. Seven Bridges Press.
- Cummins, R. (2002b): *Comments on Smith on Cummins*. in Clapin, H.: *Philosophy of Mental Representation*. Oxford: Clarendon Press.
- Cummins, R.; Blackmon, J.; Byrd, D.; Poirier, P.; Roth, M.; Schwarz, G.(2001): *Systematicity and the Cognition of Structured Domains*. in *Journal of Philosophy* Vol. 98.
- Cummins, R.; Poirier, P. (forthcoming): *Representation and Indication*. in Clapin, H.; Staines, P.; Slezak, P. (Hrsg.): *Representation in Mind: New Approaches to Mental Representation*. Westport, Conn.: Praeger.

- Danto, A. C. (1969): *Semantical Vehicles, Understanding, and Innate Ideas*. in Hook, S. (Hrsg.): *Language and Philosophy*. New York: University Press.
- Dretske, F. (1981): *Knowledge and the Flow of Information*. Cambridge, Mass.: MIT Press.
- Dudel, J.; Menzel, R.; Schmidt, R. F. (1996): *Neurowissenschaft: vom Molekül zur Kognition*. Berlin; Heidelberg; New York: Springer.
- Edelman, G. (1987): *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books.
- Elman, J.L.; Bates, E.; Johnson, M. H.; Karmiloff-Smith, A.; Parisi, D.; Plunkett, K. (1996): *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge: MIT Press.
- Fodor, J. A. (1975): *The Language of Thought*. New York: Crowell.
- Fodor, J. A. (1983): *The Modularity of Mind*. Cambridge, Mass.: MIT Press.
- Fodor, J. A. (1987): *Psychosemantics. The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass.: MIT Press.
- Fodor, J. A. (1998): *Concepts: Where cognitive science went wrong*. New York, Oxford: University Press.
- Fodor, J. A.; Lepore, E. (1992): *Paul Churchland: State Space Semantics*. in Fodor, J. A.; Lepore, E. (Hrsg.): *Holism: A Shopper's Guide*. Cambridge, Mass.: Blackwell.
- Fodor, J. A.; Lepore, E. (1996): *Reply to Churchland*. in McCauley, R. N. (Hrsg.): *The Churchlands and their Critics*. Cambridge, Mass.: Blackwell.
- Fodor, J.A.; McLaughlin, B.P. (1990): *Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work*. in *Cognition* Vol. 35.
- Fodor, J. A.; Pylyshyn, Z.(1988): *Connectionism and Cognitive Architecture: A Critical Analysis in Cognition*. Vol. 28.
- Gazzaniga, M. S.; Ivry, R. B.; Mangun, G. R. (2002): *Cognitive Neuroscience: the Biology of the Mind*. New York: W. W. Norton & Company; Zweite Auflage.
- Gilbert, C. D. (1994): *Learning: Neuronal Dynamics and Perceptual Learning*. in *Current Biology*. Vol. 4.
- Johnson, K. (forthcoming): *On the Systematicity of Language*. in *Journal of Philosophy*. Vol. 101.
- Kolb, B.; Whishaw, I.Q. (1996): *Neuropsychologie*. Heidelberg; Berlin; Oxford: Spektrum; Zweite Auflage.
- Kosslyn, S. (1980): *Image and Mind*. Cambridge, Mass.: Harvard University Press.

McGinn, C. (1989): *Mental Content*. Oxford: Blackwell.

Rupert, R.D.(2001): *Coining Terms in the Language of Thought: Innateness, Emergence, and the LOT of Cummins` s Argument against the Causal Theory of mental Content*. in *Journal of Philosophy* Vol. 98.

Samuels, R.(1998): *What Brains Won` t Tell Us about the Mind: A Critique of the Neurobiological Argument against Representational Nativism*. in *Mind and Language*. Vol. 13.

Schoups, A. (2002): *Electrophysiological Correlates of Perceptual Learning*. in Fahle, M.; Poggio, T. (Hrsg.): *Perceptual Learning*. Cambridge, Mass.: MIT Press.

Smolensky, P. (1990): *Tensor Product Variable Binding and the Representation of Symbolic Structures in Connectionist Systems*. in *Artificial Intelligence*. Vol. 46.

Thelen, E.; Smith, L. (1994): *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, Mass.: MIT Press.