

*Simulation von Wortassoziationen
mit Hilfe von mathematischen
Lernmodellen in der Psychologie*

Dissertation

Zur Erlangung des akademischen Grades eines Doktors der
Philosophie (Dr. phil.)

vorgelegt der Fakultät für Kulturwissenschaften der
Universität Paderborn

von Petra Seidensticker

Erstgutachter: Prof. Manfred Wettler
Zweitgutachter: Prof. Peter Sedlmeier

Paderborn, den 25. September 2006

Vorwort

Diese Arbeit entstand während meiner Tätigkeit als wissenschaftliche Mitarbeiterin am Lehrstuhl für Kognitive Psychologie bei Prof. Wettler an der Universität Paderborn. Ich möchte Herrn Prof. Wettler an dieser Stelle für die geleistete Betreuungsarbeit danken.

Mein Dank gilt auch Anja Austermann, die mir die englischen Assoziationsnormen für die Simulationen aufbereitet hat.

Inhaltsverzeichnis

1	Einführung	8
2	Wortassoziationen	12
3	Korpusbasierte Simulationen	17
3.1	Fenstertechnik	18
3.2	Spence & Owens	19
3.3	Assoziatives Wortnetz - Wettler & Rapp	20
3.4	HAL Hyperspace Analogue to Language - Burgess	25
3.5	LSA Latent Semantic Analysis - Landauer	28
3.6	Topic Modell - Griffiths & Steyvers	33
3.7	ICAN Modell - Lemaire & Denhiere	37
3.8	Fazit	41
4	Lerntheorien	43
4.1	Bedingter Reflex	44
4.2	Das Estes Modell - Stimulus Sampling Theorie	45
4.3	Das Rescorla & Wagner Modell	50
4.3.1	Blocking	55
4.3.2	Overshadowing (Überschattung von Reizen)	55
4.3.3	Inhibition	56
4.4	Das Gallistel Modell - Rate Estimation Theorie (RET)	56
4.4.1	Background Conditioning	60
4.4.2	Blocking	61
4.4.3	Overshadowing (Überschattung von Reizen)	63
5	Assoziieren mit den Lerntheorien	64
5.1	Semantische Konditionierung	64
5.2	Operantes Lernen und Wortassoziationen	66
5.3	Algorithmen für das Erlernen von Wortassoziationen	67
5.3.1	Algorithmus für das Estes Modell	68

<i>INHALTSVERZEICHNIS</i>	3
5.3.2 Algorithmus für das Rescorla & Wagner Modell	75
5.3.3 Algorithmus für das Gallistel Modell	79
6 Simulationen	85
6.1 Korpora	86
6.2 Lemmatisierung	87
6.3 Kriterien	89
6.4 Parameterschätzung	93
6.4.1 Worthäufigkeitseffekt	94
6.4.2 Hill-Climbing Algorithmus	97
6.4.3 Estes Modell	97
6.4.4 Rescorla & Wagner Modell	112
6.4.5 Gallistel Modell	127
7 Ergebnisse	136
7.1 Kriterium 1	142
7.2 Kriterium 2	145
7.3 Kriterium 3	147
7.4 Kriterium 4	149
7.5 Vergleich und Interpretation der Simulationsergebnisse	151
7.6 Fazit	152
8 Englische Wortassoziationen	153
9 Zusammenfassung und Interpretation	164

Abbildungsverzeichnis

6.1	Estes Modell: Ergebnisse in Bezug auf das Kriterium 1	109
6.2	Estes Modell: Ergebnisse in Bezug auf das Kriterium 2	110
6.3	Rescorla & Wagner Modell: Ergebnisse in Bezug auf das Kriterium 1	124
6.4	Rescorla & Wagner Modell: Ergebnisse in Bezug auf das Kriterium 2	125
6.5	Modellvergleich Gallistel Modell und Rescorla & Wagner Modell in Bezug auf das Kriterium 1	133
6.6	Modellvergleich Gallistel Modell und Rescorla & Wagner Modell in Bezug auf das Kriterium 2	134
7.1	Anzahl richtig vorausgesagter Primärantworten, die mit den unterschiedlichen Simulationen der einzelnen Modelle in Abhängigkeit von dem verwendeten Textkorpus erzielt wurden.	144
7.2	Anzahl Nennungen der Teilnehmer, die mit den unterschiedlichen Simulationen der einzelnen Modelle in Abhängigkeit von dem verwendeten Textkorpus vorhergesagt wurden.	146
7.3	Anzahl richtig vorausgesagter Primärantworten, die mit den unterschiedlichen Simulationen der einzelnen Modelle in Abhängigkeit von dem verwendeten Textkorpus auf Rang 1 oder Rang 2 vorhergesagt wurden.	148
7.4	Anzahl richtig vorausgesagter Primärantworten, die mit den unterschiedlichen Simulationen der einzelnen Modelle in Abhängigkeit von dem verwendeten Textkorpus auf Rang 1 bis Rang 5 vorhergesagt wurden.	150

Tabellenverzeichnis

2.1	Assoziationen zu dem Stimuluswort <i>Tisch</i>	13
2.2	Liste mit den 100 Stimuluswörtern, entnommen aus (Russell, 1970).	14
3.1	Vergleich: ICAN Modell, LSA Modell und Assoziationsnormen	40
4.1	Kontingenzmatrix für einen CS und einen US	52
4.2	Vier-Felder Tafel für kombinierten Reiz AX und Reiz X allein	53
6.1	Berechnete Assoziationen zu dem Stimuluswort <i>essen</i>	91
6.2	Assoziationen der Teilnehmer zu dem Stimuluswort <i>essen</i>	92
6.3	Estes Modell: Ermittlung unterschiedlicher Startwerte für den Hill-Climbing Algorithmus	98
6.4	Estes Modell: Erzielte Ergebnisse mit dem Hill-Climbing Algorithmus, wenn $5E - 5$ als Startwert gewählt wurde	99
6.5	Estes Modell: Erzielte Ergebnisse mit dem Hill-Climbing Algorithmus, wenn ein Startwert von $5E - 6$ verwendet wurde.	100
6.6	Estes Modell: Berechnete Assoziationen zu Frau	101
6.7	Estes Modell: Ergebnisse in Bezug auf die vier Kriterien bei Verwendung eines variablen Lernparameters	102
6.8	Estes Modell: Ergebnisse, die mit dem Parameter $\theta_{i,j} = F \cdot (I_i \cdot I_j)^2$ erzielt wurden.	104
6.9	Estes Modell: Ergebnisse, die mit dem Parameter $\theta_{i,j} = F \cdot (I_i \cdot I_j)^3$ erzielt wurden.	104
6.10	Estes Modell: Ergebnisse, die mit dem Parameter $\theta_{i,j} = F \cdot (I_i \cdot I_j)^4$ erzielt wurden.	105
6.11	Estes Modell: Ergebnisse, die mit dem Parameter $\theta_{i,j} = F \cdot (I_i \cdot I_j)^5$ erzielt wurden.	106
6.12	Estes Modell: Ergebnisse, die mit dem Parameter $\theta_{i,j} = F \cdot (I_i \cdot I_j)^6$ erzielt wurden.	107

6.13	Estes Modell: Assoziationen zu Frau, die mit dem Parameter $\theta_{i,j} = 5E - 20 \cdot (I_i \cdot I_j)^6$ berechnet wurden.	111
6.14	Rescorla & Wagner Modell: Ergebnisse, die mit dem Parameter $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^5$ erzielt wurden, in Abhängigkeit von den unterschiedlichen Faktoren.	114
6.15	Rescorla & Wagner Modell: Ergebnisse, die mit dem Parameter $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^6$ erzielt wurden, in Abhängigkeit von den unterschiedlichen Faktoren.	115
6.16	Rescorla & Wagner Modell: Ergebnisse, die mit unterschiedlichen Parametern für Lernen und Vergessen erzielt wurden. Lernparameter: $\alpha_i\beta_j = L \cdot (I_i \cdot I_j)^5$, Parameter für Vergessen: $\alpha_i\beta_j = V \cdot (I_i \cdot I_j)^5$	116
6.17	Rescorla & Wagner Modell: Ergebnisse, die mit unterschiedlichen Parametern für Lernen und Vergessen erzielt wurden. Lernparameter: $\alpha_i\beta_j = L \cdot (I_i \cdot I_j)^6$, Parameter für Vergessen: $\alpha_i\beta_j = V \cdot (I_i \cdot I_j)^6$	117
6.18	Rescorla & Wagner Modell: Ergebnisse, die mit unterschiedlichen Parametern für Lernen und Vergessen erzielt wurden. Lernparameter: $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^L$, Parameter für Vergessen: $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^V$	118
6.19	Rescorla & Wagner Modell ohne Regel für Vergessen: Ergebnisse, die mit dem Lernparameter $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^4$ erzielt wurden.	119
6.20	Rescorla & Wagner Modell ohne Regel für Vergessen: Ergebnisse, die mit dem Lernparameter $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^5$ erzielt wurden.	119
6.21	Rescorla & Wagner Modell ohne Regel für Vergessen: Ergebnisse, die mit dem Lernparameter $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^6$ erzielt wurden.	120
6.22	Rescorla & Wagner Modell ohne Regel für Vergessen: Ergebnisse, die mit dem Lernparameter $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^{5.5}$ erzielt wurden.	120
6.23	Rescorla & Wagner Modell: Ergebnisse, die mit dem Lernparameter $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^5$ und einer Rotation des Fensterinhaltes erzielt wurden.	121
6.24	Rescorla & Wagner Modell: Ergebnisse, die mit dem Lernparameter $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^6$ und einer Rotation des Fensterinhaltes erzielt wurden.	122
6.25	Gallistel Modell: Ergebnisse der Simulationen mit Gewichtung	128

6.26	Gallistel Modell: Ergebnisse der Simulationen mit Gewichtung, wenn Potenzierungen zwischen 4.2 und 6.2 verwendet wurden.	129
6.27	Gallistel Modell: Ungewichtete Assoziationen zu dem Stimuluswort <i>essen</i>	130
6.28	Gallistel Modell: Gewichtete Assoziationen zu dem Stimuluswort <i>essen</i>	131
6.29	Gallistel Modell: Ungewichtete Assoziationen zu dem Stimuluswort <i>Kind</i>	132
6.30	Gewichtete Assoziationen zu dem Stimuluswort <i>Kind</i>	135
7.1	Simulationen mit dem Textkorpus SZ95	139
7.2	Simulationen mit dem Textkorpus SZ96	140
7.3	Simulationen mit dem Textkorpus SZ97	141
7.4	Simulationen mit dem Textkorpus SZ	142
8.1	Simulation von englischen Wortassoziationen mit dem Estes Modell, wobei ein konstanter Lernparameter θ für alle Wortpaare verwendet wurde.	155
8.2	Simulation von englischen Wortassoziationen mit dem Estes Modell mit $\theta_{i,j} = F \cdot (I_i \cdot I_j)$	160
8.3	Simulation von englischen Wortassoziationen mit dem Estes Modell mit $\theta_{i,j} = F \cdot (I_i \cdot I_j)^2$	161
8.4	Simulation von englischen Wortassoziationen mit dem Estes Modell mit $\theta_{i,j} = F \cdot (I_i \cdot I_j)^3$	162
8.5	Simulation von englischen Wortassoziationen mit dem Estes Modell mit $\theta_{i,j} = F \cdot (I_i \cdot I_j)^4$	163

Kapitel 1

Einführung

In dieser Arbeit soll untersucht werden, ob Wortassoziationen mit Hilfe von mathematischen Lernmodellen durch eine Computersimulation vorausgesagt werden können. Hierfür wurden drei mathematische Lerntheorien, die jeweils ein mathematisches Lernmodell enthalten, ausgewählt. Dies waren die Stimulusauswahltheorie von Estes (Estes, 1950), die Theorie von Rescorla & Wagner (Wagner & Rescorla, 1972) und die Rate Estimation Theory (RET) von Gallistel (Gallistel, 1990).

Wortassoziationen haben in der Psychologie schon immer eine wichtige Rolle gespielt. Sie steuern kognitive Prozesse, wie z.B. Lernen, Erinnern und Kreativität (Sarnoff, 1962). Sie wurden zum ersten mal von Francis Galton (Galton, 1880) untersucht.

Wortassoziationen werden in einem Assoziationsversuch ermittelt. In einem Assoziationsversuch erhalten die Teilnehmer ein Wort (Stimuluswort) auf das sie mit einem anderem Wort (Responsewort), das ihnen als erstes einfällt, antworten sollen. Die Produktion der Responsewörter, die die Teilnehmer im Assoziationsversuch nennen, wird in der Assoziationstheorie durch Kontiguitätslernen erklärt (Wettler, Rapp & Ferber, 1993). William James (1890) (James, 1890) beschreibt das Kontiguitätsprinzip folgendermaßen:

“Objects once experienced together tend to become associated in the imagination, so that when any one of them thought of, the others are likely to be thought also, in the same order of sequence or coexistence as before. This statement we may name the law of mental association by contiguity.” (James, 1890, p. 561)

In der heutigen Psychologie werden unterschiedliche Ansichten darüber vertreten, ob das Kontiguitätsprinzip allein ausreicht, um Wortassoziationen zu erklären. Matthäus (Matthäus, 1980) kritisiert die Interpretation des As-

soziationsexperimentes von Galton, indem er behauptet, dass die im Assoziationsexperiment ermittelten Beziehungen zwischen Stimuluswörtern und Responsewörtern nicht außerhalb der experimentellen Situation zu beobachten sind. Deshalb eignet sich der Assoziationsversuch laut Matthäus nicht zur Untersuchung von sprachlichen Prozessen.

Die freien Wortassoziationen sind nach Clark (Clark, 1970) nicht die Folge von Kontiguitätslernen, sondern die Folge von einer Serie von symbolischen informationsverarbeitenden Prozessen. Nach Clark wird das Stimuluswort zuerst semantisch encodiert. Diese semantische Repräsentation wird dann durch die Änderung von einzelnen Merkmalen zu einer neuen semantischen Repräsentation transformiert. Das Responsewort, das am besten mit dieser neuen semantischen Repräsentation übereinstimmt, würde dann als assoziative Antwort gegeben.

Die Kontiguitätstheorie wurde häufig von kognitiven Psychologen abgelehnt, weil sie viele Charakteristika des sprachlichen Verhaltens nicht erklären kann (Lashley, 1951). Für die Kontiguitätstheorie sprechen mehrere Arbeiten, in denen mit Hilfe von neuronalen Netzen sprachliche und visuelle Prozesse durch Kontiguitätslernen erfolgreich simuliert werden konnten (Kohonen, 1989). Nach Wettler (Wettler, Rapp & Ferber, 1993) soll man sich nicht zwischen zwei gegenseitig sich ausschließenden Theorien, der Kontiguitätstheorie und dem symbolischen Ansatz, entscheiden, sondern man soll untersuchen, welche Verhaltensbereiche mit den verschiedenen Theorien erklärt werden können.

Es konnte nachgewiesen werden, dass sich die im Assoziationsexperiment gefundenen Beziehungen zwischen Wörtern auch in anderen Bereichen sprachlichen Verhaltens finden lassen. In (Wettler, Rapp & Ferber, 1993) konnten die freien Wortassoziationen aufgrund des gemeinsamen Auftretens der Wörter in Texten weitgehend vorausgesagt werden. Nach Wettler et al. (Wettler, Rapp & Ferber, 1993) sind Wortassoziationen Ergebnisse von assoziativen Lernprozessen. Die Assoziationsstärken korrelieren mit der Häufigkeit des gemeinsamen Auftretens (Wettler, Rapp & Ferber, 1993).

In dieser Arbeit sollen die Assoziationsstärken nicht durch das Zählen von Kookkurrenzen gewonnen werden, sondern durch die Simulation von mathematischen Lernmodellen. Die auf der Grundlage von mathematischen Lernmodellen gewonnenen Assoziationen sollen anhand von menschlichen Assoziationen überprüft werden. Später sollen mit Hilfe des mathematischen Lernmodells, das bei der Voraussage der deutschen Assoziationen am besten abschneidet, noch englische Assoziationen vorausgesagt werden.

Die mathematischen Lernmodelle können grob in drei Klassen eingeteilt werden. Dies sind einmal die Kontiguitätsmodelle, zu denen das Modell von Estes gehört. Sie zählen historisch gesehen zu den ältesten Lernmodellen.

Danach wurden die Kontingenzmodelle, zu denen das Rescorla & Wagner Modell gehört, entwickelt. Später folgten die Ratio - Zeit Modelle, zu denen das Gallistel Modell gehört.

Zum Schluss sollen die Assoziationen, die mit den drei Lernmodellen ermittelt wurden, untereinander verglichen werden. Mit Hilfe dieses Vergleichs soll die folgende Frage beantwortet werden:

Welche Lerntheorie eignet sich am besten zur Voraussage von menschlichen Wortassoziationen?

Die Arbeit ist in neun Kapitel aufgegliedert. Im zweiten Kapitel wird näher auf das Thema Wortassoziationen und die verwendeten Assoziationsnormen eingegangen. Der oben erwähnte Assoziationsversuch wird genau erläutert. Danach wird beschrieben, welche Relevanz die Wortassoziationen in der Psychologie haben.

Das dritte Kapitel beschäftigt sich mit korpusbasierten Simulationen von sprachlichen Leistungen. Es werden die Arbeiten von Spence & Owens (Spence & Owens, 1990), Wettler & Rapp (Wettler, Rapp & Ferber, 1993), Lemaire & Denhiere (Lemaire & Deniere, 2004) und Griffiths & Steyvers (Griffiths & Steyvers, 2002) zur Berechnung von Wortassoziationen und die Arbeiten von Burgess (Burgess, 1998) und Landauer (Landauer & Dumais, 1997) zur Berechnung von Wortähnlichkeiten vorgestellt.

Im vierten Kapitel werden die einzelnen mathematischen Lerntheorien erläutert und deren Bedeutung in der Psychologie diskutiert. Ausgehend vom Assoziationsgesetz wird die klassische Konditionierung beschrieben. Die beiden Modelle, das Modell von Rescorla & Wagner und das Gallistel Modell, wurden speziell für die Beschreibung der klassischen Konditionierung entwickelt. Die Stimulus Sampling Theorie nach Estes erklärt das Lernen von Wahrscheinlichkeiten.

Im fünften Kapitel wird die Anwendung der Lerntheorien auf das Erlernen von Wortassoziationen beschrieben. Es wird beschrieben, wie Staats das Erlernen der Bedeutung von Wörtern teilweise auf eine Konditionierung höherer Ordnung und teilweise auf operantes Konditionieren zurückgeführt hat. Beide Konditionierungsarten können mit den drei Lerntheorien erklärt werden. Es wird beschrieben, wie die Wortassoziationen nach den drei Lerntheorien gelernt werden können. Für alle drei Lernmodelle wird ein Algorithmus angegeben, der zeigt, wie mit dem jeweiligen Modell Wortassoziationen gelernt werden.

Das sechste Kapitel beschreibt die Simulationen, die mit den drei Lernmodellen durchgeführt wurden. Es werden die verwendeten Textkorpora und deren Aufbereitung für die Berechnung von Wortassoziationen vorgestellt. Zur Beurteilung der berechneten Assoziationen wurden vier Gütekriterien aufgestellt, die in dem Kapitel Simulationen beschrieben werden. Da die

drei Lernmodelle freie Parameter enthalten, wurde für jedes Lernmodell eine Parameterschätzung anhand eines Kalibrierkorpus vorgenommen. Später wurden die Simulationen anhand eines Testkorpus durchgeführt, um die Allgemeingültigkeit der Parameter, die bei der Parameterschätzung gewonnen wurden, zu überprüfen. Das Kalibrierkorpus diente nur zur Parameterschätzung, während der Vergleich der Simulationsergebnisse am Testkorpus vorgenommen wurde.

Im siebten Kapitel werden zunächst die Ergebnisse der Simulationen der einzelnen Lernmodelle vorgestellt. Anschließend werden die Ergebnisse der Simulationen miteinander verglichen, um die Frage zu beantworten, welche Theorie kann die Wortassoziationen der Teilnehmer am besten voraussagen.

Im achten Kapitel wird die Berechnung von englischen Wortassoziationen beschrieben. Für die Voraussage der englischen Wortassoziationen wird das Modell, das bei der Voraussage der deutschen Wortassoziationen am besten abgeschnitten hat, verwendet. Die berechneten englischen Wortassoziationen wurden anhand der Assoziationsnormen, die von Nelson, McEvoy und Schreiber (Nelson, McEvoy & Schreiber, 1999) erhoben wurden, überprüft.

Im neunten Kapitel folgt eine Zusammenfassung und eine Interpretation der Ergebnisse.

Kapitel 2

Wortassoziationen

In der allgemeinen Psychologie werden freie Wortassoziationen erhoben, um die Stärke der Verwandtschaft zwischen Wörtern bestimmen zu können (Wettler, 1980). In dem dazu durchgeführten Assoziationsexperiment haben die Teilnehmer die Aufgabe, zu einem vorgegebenen Stimuluswort, das erste Wort, das ihnen einfällt, zu nennen. Wird dieses Assoziationsexperiment mit vielen Teilnehmer durchgeführt, so lässt sich aus den Responsewörtern der Teilnehmer zu jedem Stimuluswort eine Häufigkeitstabelle erstellen. Diese Häufigkeitstabellen, die zu verschiedenen Stimuluswörtern erstellt werden, werden Assoziationsnormen genannt. Für das Stimuluswort *Tisch* ist eine solche Häufigkeitstabelle in der Tabelle 2.1 dargestellt. Das Responsewort, das von den meisten Teilnehmern genannt wird, wird Primärantwort genannt. Zu dem Stimuluswort *Tisch* ist die Primärantwort *Stuhl*. Sie wurde von 97 Teilnehmern als Responsewort genannt.

Die bekanntesten Assoziationsnormen wurden von Kent und Rosanoff (1910) erhoben. Sie verwendeten 100 englische Reizwörter. In den 50er Jahren hatten Russell und Meseck deutschsprachige Assoziationsnormen erhoben, die aber erst 1970 von Russell veröffentlicht wurden. Die Assoziationen zu dem Stimuluswort *Tisch* in der Tabelle 2.1 sind aus den Assoziationsnormen, die von Russell (Russell, 1970) veröffentlicht wurden, entnommen. Russell und Meseck verwendeten die deutschen Übersetzungen, der von Kent und Rosanoff verwendeten englischen 100 Reizwörter, als Reizwörter. Die Teilnehmer erhielten jeweils eine Liste mit diesen 100 Reizwörtern. Diese Liste bestand aus zwei Seiten. Die Stimuluswörter waren jeweils in zwei Spalten auf einer Seite aufgelistet. Eine Spalte enthielt jeweils 25 Stimuluswörter. Die Stimuluswörter waren von 1 bis 100 durchnummeriert. Hinter jedem Stimuluswort war eine Linie auf der die Teilnehmer das Responsewort schreiben sollten. Die Teilnehmer bekamen die beiden Seiten mit den Stimuluswörtern so ausgehändigt, dass sie die Stimuluswörter nicht schon vor Versuchsbeginn

Assoziation	Anzahl
Stuhl	97
Bein	49
Tuch	23
essen	22
Decke	20
Bank	8
Holz	8
Platte	8
Tischtuch	8
Möbel	5
Tischbein	5

Tabelle 2.1: Assoziationen der Teilnehmer zu dem Stimuluswort *Tisch*. In der rechten Spalte sind die Anzahl der Teilnehmer aufgeführt, die das jeweilige Responsewort im Assoziationsversuch genannt hatten. In der linken Spalte sind Responsewörter aufgeführt, die von mindestens fünf Teilnehmern im Assoziationsversuch genannt wurden.

sehen konnten. Berichte über die Instruktionen an die Teilnehmer sind in der Literatur häufig ungenau, deswegen sind die genauen Anweisungen aus (Russell, 1970) übernommen worden. Sie liegen aber nur in Englisch vor. Die Teilnehmer erhielten eine deutsche Übersetzung des folgenden Textes:

“At the University of Wuerzburg, a series of investigations of verbal behavior is being carried out. The following experiment is one of this series. It is on free association.

You have received two sheets. Please write your name on the outside of the first sheet and answer the questions listed before you. Open the sheet only after I have read you the instructions.

On the inside of these sheets you will find 100 stimulus words. Your task is to write after each word the first word you think of. Write only one word for each response. Begin with word number 1, read it and write beside it the first word that occurs to you; then read the second word, and so on, until the list is finished. Work as fast as you can and do not leave any words out.

When you have finished, close the sheets and write back the letter which you see written on the board at that time. (Demonstration) Are there any questions? Are you ready? Please begin!” (Russell 1970, p. 55)

Nach diesen Anweisungen fingen die Teilnehmer mit der Abarbeitung der Wörter an. Vier Minuten später wurde der Buchstabe A an die Tafel geschrieben. Danach folgte alle 30 Sekunden ein neuer Buchstabe, der an die Tafel geschrieben wurde. Die Buchstaben erschienen in alphabetischer Reihenfolge an der Tafel. Durch das Anschreiben der Buchstaben sollten die Teilnehmer motiviert werden, schnell zu arbeiten. Wenn sie alle Reizwörter bearbeitet hatten, sollten sie den aktuellen Buchstaben an der Tafel in den Antwortbogen eintragen. Dadurch erhielten die Versuchsdurchführer einen ungefähren Überblick, wie lange die einzelnen Teilnehmer für die Antworten auf die 100 Reizwörter benötigt hatten.

Die Teilnehmer dieses Assoziationsversuchs waren Schüler (Gymnasium und Oberrealschule) aus Würzburg, Aschaffenberg und Schweinfurt und Psychologiestudenten aus Würzburg. Insgesamt hatten 331 Personen im Alter von 16 bis 30 Jahren an diesem Assoziationsversuch teilgenommen.

In der Tabelle 2.2 sind die 100 Stimuluswörter aufgelistet.

Tisch	dunkel	Musik	Krankheit	tief
weich	essen	Berg	schwarz	Bequemlichkeit
Hand	kurz	Obst	Schmetterling	glatt
kommandieren	Stuhl	süß	pfeifen	kalt
langsam	Haus	Frau	Mann	wünschen
Fluss	schön	Fenster	rauh	Fuß
Spinne	Nadel	rot	schlafen	Zorn
Teppich	Mädchen	hoch	arbeiten	sauer
Erde	Sorge	Soldat	hart	Adler
Magen	Stiel	Lampe	träumen	gelb
Brot	Gerechtigkeit	Junge	Licht	Gesundheit
Bibel	Gedächtnis	Schaf	Bad	schnell
blau	hungrig	Priester	Ozean	Kopf
Ofen	lang	Religion	Whisky	Kind
bitter	Hammer	durstig	Stadt	Quadrat
Butter	Arzt	laut	Dieb	Löwe
Freude	Bett	schwer	Tabak	Baby
Mond	Schere	ruhig	grün	Salz
Straße	König	Käse	Blüte	ängstlich
Bürger	weiß	Kohl	Häuschen	Hammelfleisch

Tabelle 2.2: Liste mit den 100 Stimuluswörtern, entnommen aus (Russell, 1970).

Eine viel größere Sammlung von Assoziationsnormen gibt es für die engli-

sche Sprache. Sie wurden von Nelson, McEvoy und Schreiber (Nelson, McEvoy & Schreiber, 1999) für insgesamt 5019 Stimuluswörter erhoben. An dieser Erhebung hatten 6000 Personen aus der gesamten USA teilgenommen. Wegen der großen Anzahl von Stimuluswörtern konnte nicht jeder Teilnehmer der Erhebung alle 5019 Stimuluswörter bearbeiten. Ein Teilnehmer erhielt 100 bis 120 Stimuluswörter. Zu jedem Stimuluswort wurden ca. 150 Teilnehmer befragt. Die Teilnehmer erhielten die Aufgabe, jeweils das erste Wort, das ihnen zu den Stimuluswörtern einfiel, aufzuschreiben. Pro Stimuluswort durfte nur ein Responsewort aufgeschrieben werden. Die Stimuluswörter in der Erhebung von Nelson, McEvoy und Schreiber (Nelson, McEvoy & Schreiber, 1999) bestehen hauptsächlich aus Substantiven. Der Anteil der Substantive in dieser Erhebung beträgt 76 Prozent. Nur 7 Prozent der Stimuluswörter sind Verben und 13 Prozent sind Adjektive. Die Assoziationsnormen sind im Internet unter der Adresse <http://www.usf.edu/FreeAssociation> erhältlich.

In einer großen Zahl von Experimenten konnte gezeigt werden, dass die Wahrnehmung, dass Erlernen und dass Erinnern von sprachlichem Material im hohen Maße durch Wortassoziationen beeinflusst wird. Zwei dieser Experimente sollen im folgenden kurz beschrieben werden.

Deese untersuchte Intrusionen bei der freien Reproduktion von Wortlisten (Deese, 1959b). Dies sind Wörter, die von den Teilnehmern reproduziert werden, obwohl sie nicht in der Lernliste enthalten sind. Die Teilnehmer erhielten Listen mit 15 stark assoziierten Wörtern zu einem Wort, das eine Art Köder (eng. lure) darstellte, das selber nicht auf der Liste stand. Ein Beispiel für eine Liste mit stark assoziierten Wörtern zu dem Wort *Schlaf* ist die folgende Liste mit den Wörtern: *Bett, Traum, müde, wach* usw. . Bei der freien Reproduktion dieser Liste zeigte sich, dass die sogenannten Köder von über 40 Prozent der Teilnehmer falsch erinnert wurden. Bei der Liste mit den assoziierten Wörtern zu *Schlaf* hatten 61 Prozent der Teilnehmer das Wort *Schlaf* erinnert, obwohl es gar nicht auf der Liste stand. Andere Wörter, die als Köder in dem Experiment dienten, wurden niemals falsch erinnert. Deese konnte die Intrusionsraten der Wörter, die als Köder dienten, mit Hilfe von Wortassoziationen voraussagen. Als Maß für die Störungsraten verwendete er den Durchschnitt der Assoziationsstärken zwischen den Wörtern auf der Liste und dem Wort, das als Köder diente. Er erhielt eine Korrelation von $r=0.8$ zwischen vorausgesagten Störungen und tatsächlichen Störungen.

Der Einfluss von Wortassoziationen konnte auch in Bahnungsexperimenten (Primingexperimente) nachgewiesen werden. In einem typischen Priming - Experiment sehen die Teilnehmer eine Zeichenkette, die Zielwort genannt wird. Die Teilnehmer sollen entscheiden, ob die dargebotene Zeichenkette ein Wort darstellt oder nicht. Kurz bevor die Teilnehmer die Zeichenkette sehen, sehen die Teilnehmer noch ein anderes Wort, das sich Bahnungs- oder

Primewort nennt. In den Bahnungsexperimenten zeigte sich, dass das Zielwort umso schneller erkannt wurde, je stärker die Assoziationsstärke zwischen Bahnungs- und Zielwort war.

Kapitel 3

Korpusbasierte Simulationen von Wortähnlichkeiten und Wortassoziationen

In diesem Kapitel sollen eine Reihe von Arbeiten beschrieben werden, in denen die Verbundenheit oder Ähnlichkeit von Wörtern aufgrund der Häufigkeit ihres gemeinsamen Auftretens berechnet wurde. Diese Arbeiten sind für unser Thema insofern relevant, als das gemeinsame Auftreten von Wörtern beim Erlernen von Wortassoziationen die entscheidende Variable darstellt.

Um das gemeinsame Auftreten von Wörtern bestimmen zu können, muss ein Bereich festgelegt werden innerhalb dessen zwei Wörter als Wortpaare betrachtet werden sollen. Eine Möglichkeit den Bereich des gemeinsamen Auftretens von Wörtern festzulegen, ist die sogenannte Fenstertechnik, die im nächsten Abschnitt erläutert wird. Die Fenstertechnik ist eine Möglichkeit die Häufigkeit des gemeinsamen Auftretens und damit auch den gemeinsamen Kontext von Wörtern zu bestimmen.

Eine weitere Möglichkeit der Festlegung des gemeinsamen Kontextes von Wörtern kommt aus dem Information Retrieval. Der gemeinsame Kontext wird hier entweder durch ein ganzes Dokument oder durch einen Abschnitt festgelegt. Für jedes Dokument wird ein Vektor gebildet, der die Häufigkeiten der Wörter, die in diesem Dokument vorkommen, enthält. Nachdem für jedes Dokument ein solcher Vektor gebildet worden ist, erhält man eine *Term x Dokument* oder *Wort x Dokument* Matrix. Bei der Fenstertechnik erhält man eine *Wort x Wort* Matrix.

3.1 Fenstertechnik

Bei der Verwendung der Fenstertechnik wird ein Fenster mit konstanter Anzahl von Wörtern eingelesen. Dieses Fenster wird wortweise über einen Text geschoben, indem rechts im Fenster das neue Wort eingefügt wird und das Wort am linken Rand aus dem Fenster entfernt wird. Dieser Vorgang wird solange wiederholt, bis der ganze Text verarbeitet worden ist.

Die Fenstertechnik wird an einem Wortfenster, das aus fünf Wörtern besteht, demonstriert:

Das Wortfenster wird über den Textkorporus geschoben.
 Das Wortfenster wird über den Textkorporus geschoben.
 Das Wortfenster wird über den Textkorporus geschoben.

Die Idee hinter dieser Methode ist, dass Wörter, die in der gleichen Umgebung miteinander vorkommen, auch miteinander verwandt sind.

Um die Verwandtschaft der Wörter zu bestimmen, müssen mit den Wörtern im Fenster Operationen durchgeführt werden. Eine solche Operation ist im einfachsten Fall nur ein Zählen von Häufigkeiten. Es wird gezählt, wie häufig kommt ein Wortpaar gemeinsam im Fenster vor. Bei der Verwendung von komplexeren Operationen werden Gleichungen verwendet, die inkrementell die Stärke der Verbindung zwischen zwei Wörtern verändern. Da die Operationen auf Wortpaare im Fenster ausgeführt werden, muss festgelegt werden, wie Wortpaare im Fenster gebildet werden. Bei einer Fensterlänge von 2 besteht das Fenster nur aus zwei Wörtern, und es existiert nur eine Möglichkeit, ein Wortpaar zu bilden. Ab einer Fensterlänge von drei gibt es zwei unterschiedliche Varianten, Wortpaare im Fenster zu bilden:

1. Variante: Jedes Wort bildet mit jedem anderen Wort im Fenster ein Wortpaar, d.h. alle n Worte im Fenster werden kreuzweise miteinander verknüpft. Dies führt bei einer Fensterlänge von n zur Bildung von $\frac{n(n-1)}{2}$ Wortpaaren.
2. Variante: Es wird ein Wort im Fenster ausgewählt, und es werden nur Wortpaare von Wörtern im Fenster zu dem ausgewählten Wort gebildet. Dies führt zur Bildung von $n - 1$ Wortpaaren.

Bei der 1. Variante der Wortpaarbildung werden Wörter, die nur einen geringen Abstand zueinander haben, mehrfach betrachtet, da sie über mehrere Fensterverschiebungen gemeinsam im Fenster verbleiben. Direkte Wortnachbarn im Fenster werden bei einer Fensterlänge von n genau $n - 1$ mal als Wortpaar betrachtet. Bei der 2. Variante werden direkte Wortnachbarn nur

einmal als Wortpaar betrachtet. Durch die 1. Variante werden Wörter innerhalb des Fensters mit einem geringeren Abstand zueinander häufiger als Wortpaar betrachtet, als Wörter mit einem größeren Abstand zueinander. Dagegen wird bei der 2. Variante die Mehrfachbetrachtung von Wörtern mit geringen Abstand zueinander vermieden.

In den folgenden Abschnitten werden die einzelnen Arbeiten, die die Verbundenheit oder Ähnlichkeit von Wörtern berechnen, vorgestellt.

3.2 Spence & Owens

Spence & Owens (Spence & Owens, 1990) waren die Ersten, die den Zusammenhang zwischen Wortassoziationen und dem gemeinsamen Auftreten von Wörtern untersucht haben. Sie verwendeten dabei die im obigen Abschnitt beschriebene Fenstertechnik. Die Größe des Wortfensters wurde aber nicht über die Anzahl Wörter sondern über die Anzahl Buchstaben definiert. Die Wortpaare werden mit Hilfe der 1. Variante aus dem Abschnitt Fenstertechnik gebildet.

Spence & Owens nehmen an, dass die Wörter von Stimulus - Response Paaren aus den Assoziationsnormen häufiger gemeinsam vorkommen als die Wörter eines Wortpaares, das aus dem gleichen Stimuluswort aber einem nicht ähnlichen Responsewort mit gleicher Worthäufigkeit, wie das zugehörige Responsewort aus den Assoziationsnormen, besteht. Die Anzahl der Kookkurrenzen von zwei assoziierten Wörtern, wie z.B. OCEAN und WATER, sollte zum Beispiel signifikant größer sein als die Anzahl der Kookkurrenzen von OCEAN und HAND, wobei HAND ungefähr so häufig wie WATER im Textkorpus vorkam. Zur Überprüfung dieser Hypothese verwendeten Spence and Owens (Spence & Owens, 1990) den Brown - Korpus, das ca. 1 Million Wörter enthält. Die Stimulus - Response Paare wurden aus den Palermo & Jenkins (Palermo & Jenkins, 1964) Normen entnommen. Die Originalliste enthielt 200 Wörter, davon waren 80 Substantive. Von diesen 80 Substantiven wurden nur die Wörter ausgewählt, die mit einer Häufigkeit von mindestens 10 zu einer Millionen in der Sprache auftraten. Dies waren insgesamt 47 Stimuluswörter. Wenn die Primärantwort zu einem Stimuluswort kein Substantiv war, so wurde das Substantiv ausgewählt, das als häufigstes Responsewort genannt wurde. Um die Auftretenshäufigkeiten von Wortpaaren zählen zu können, wurden im Brown - Korpus zuerst alle Satzzeichen entfernt und alle Großbuchstaben in Kleinbuchstaben umgewandelt. Dann wurden für alle Stimulus - Response Paare die Kookkurrenzen in einem Bereich von 10000 Zeichen oder weniger gezählt. Es wurde auch die Distanz zwischen dem ersten Buchstaben des Stimuluswortes und dem ersten Buch-

staben des Responsewortes in Buchstaben gezählt. Ein Wort im Plural oder im Singular wurde getrennt untersucht, d.h wurde nach einem Wort im Singular gesucht, so wurde auch nur das Wort im Singular verwendet und nicht die Pluralform von diesem Wort.

Für 47 assoziierte Substantivpaare aus den Assoziationsnormen wurden insgesamt 222 Kookkurrenzen im Brown - Korpus gezählt, wenn ein Fenster von ca. 50 Wörtern (nicht mehr als 250 Buchstaben) verwendet wurde. Bei den 47 nicht assoziierten Substantivpaaren wurden dagegen nur 37 Kookkurrenzen gefunden. Somit tendierten assoziierte Wortpaare zu häufigeren gemeinsamen Auftreten als Paare von nicht assoziierten Wörtern. Die Auftretenshäufigkeit im Brown Korpus war signifikant ähnlich der Assoziationsstärke zwischen dem Stimulus und dem Responsewort. Um diese Hypothese zu überprüfen, wurden von den Kookkurrenzen von assoziierten Wörtern die Kookkurrenzen von den nicht assoziierten Wörtern subtrahiert. Diese Differenz wurde mit den Assoziationsnormen von Palermo und Jenkins (Palermo & Jenkins, 1964) korreliert. Stimulus - Response Paare mit hoher Assoziationsstärke traten häufiger gemeinsam auf als Paare mit einer geringeren Assoziationsstärke.

3.3 Assoziatives Wortnetz - Wettler & Rapp

Wettler und Rapp (Wettler, Rapp & Ferber, 1993), (Wettler, Rapp & Sedlmeier, 2005) verwendeten für die Vorhersage von freie Wortassoziationen ebenfalls die oben beschriebene Fenstertechnik. Sie hatten ihr Modell sowohl an deutschen Assoziationsnormen wie auch an englischen Assoziationsnormen überprüft. In (Wettler, Rapp & Ferber, 1993) wurden die vom Modell vorausgesagten Assoziationen mit den von Russell & Meseck erhobenen Assoziationsnormen verglichen. Dazu wurde ein Vokabular von insgesamt 65356 Wörtern verwendet. Dieses bestand einmal aus allen 2012 Wörtern, die im Assoziationsversuch als Stimulus- oder als Responsewort vorkamen, und aus Wörtern, die in den Korpora des Institutes für Deutsche Sprache mindestens zehn mal vorkamen. Die Korpora enthielten insgesamt ca. 21 Millionen Wörter.

Die Assoziationen zwischen einzelnen Wörtern werden laut Wettler und Rapp (Wettler, Rapp & Ferber, 1993), (Wettler, Rapp & Sedlmeier, 2005) nach dem Kontiguitätsgesetz gelernt. Der Berechnung der Assoziationsstärken wurden folgende Gleichungen zugrunde gelegt:

$$a_{i,j}(t + 1) = a_{i,j}(t) + \theta_1 \cdot (1 - a_{i,j}(t)) \quad (3.1)$$

$$a_{i,j}(t+1) = a_{i,j}(t) \cdot (1 - \theta_2) \quad (3.2)$$

Mit $a_{i,j}(t)$ wird die Assoziationsstärke zweier Wörter i und j zum Zeitpunkt t bezeichnet. Die Gleichung 3.1 führt zu einer Erhöhung der Assoziationsstärke, wenn das Wort i gemeinsam mit dem Wort j auftritt. Tritt nur das Wort i auf, nicht aber das Wort j , so wird die Assoziationsstärke zwischen i und j durch die Gleichung 3.2 verringert. In Gleichung 3.1 gibt θ_1 die Lernrate an, d.h. wie steil die Lernkurve verläuft. θ_2 in Gleichung 3.2 wird als Hemmungsrate bezeichnet, sie gibt an wie stark die Lernkurve abfällt. Wird angenommen, dass die Lernrate θ_1 gleich der Hemmungsrate θ_2 ist, so ist die zu erwartende Assoziationsstärke $a_{i,j}$ gleich der bedingten Wahrscheinlichkeit von j bei gegebenen i (Wettler, Rapp & Sedlmeier, 2005).

$$a_{i,j} = p(j|i) \quad (3.3)$$

Die Gleichung 3.3 lässt sich aber nur dann aus den Gleichungen 3.1 und 3.2 herleiten, wenn die Wahrscheinlichkeit, dass j bei gegebenen i nicht auftritt, gleich $1 - p(j|i)$ ist. Diese Annahme gilt aber nur bei stochastischer Unabhängigkeit (Wettler, Rapp & Sedlmeier, 2005). Die Herleitung von Gleichung 3.3 aus den Gleichungen 3.1 und 3.2 wird auch im Kapitel Lerntheorien, Abschnitt *Das Estes Modell* beschrieben. Dies sind dort die Gleichungen 4.11, 4.12 und 4.13, wenn π in diesen Gleichungen durch $p(j|i)$ ersetzt wird.

Für $\theta_1 = \theta_2$ entsprechen die obigen beiden Gleichungen den Gleichungen im Estes Modell, das im Kapitel Lerntheorien beschrieben wird. Für $\theta_1=0.5$ und $\theta_2=0.1$ entsprechen die obigen Gleichungen den Gleichungen im ICAN Modell ohne die Gleichung für die Kookkurrenzen zweiter Ordnung. Das ICAN Modell wird im letzten Abschnitt dieses Kapitels beschrieben.

Die Assoziationsstärke zweier Wörter i und j kann nach der Formel 3.3 berechnet werden, wenn vorher die Worthäufigkeiten und die Häufigkeiten von Wortpaaren (Kookkurrenzen) gezählt wurden (Wettler, Rapp & Sedlmeier, 2005).

$$a_{i,j} = \frac{h(i \cap j)}{h(j)} \quad (3.4)$$

Um die Kookkurrenzen zu ermitteln, wurde ein Fenster der Länge 25 verwendet. Bei der Wortpaarbildung wurde die 2. Variante aus dem Abschnitt Fenstertechnik verwendet, d.h. es wurden nur die Kookkurrenzen zu einem ausgewählten Wort im Fenster gezählt. Das ausgewählte Wort war das Wort in der Mitte des Fensters. Als ausgewählte Wörter wurden nur Stimuluswörter verwendet. Somit wurde für jedes Stimuluswort gezählt,

wie häufig es im Textkorpus in einer Umgebung von höchstens 12 Wörtern mit irgendeinem anderen Wort aus dem Vokabular vorkam. So entstand eine Häufigkeitstabelle für die einzelnen Wortpaare aus dem Vokabular. Die Häufigkeit des gemeinsamen Auftretens von Wort i und Wort j wird mit $h(i \cap j)$ bezeichnet, während $h(j)$ die Häufigkeit eines Wortes im gesamten Textkorpus bezeichnet. Wurden die Assoziationsstärken mit der Gleichung 3.4 berechnet, so wurden häufig Wörter mit einer sehr hohen Auftretenshäufigkeit als assoziative Antworten berechnet (Worthäufigkeitseffekt), d.h häufige Wörter erhielten eine hohe Assoziationsstärke. Deshalb wurde in (Wettler, Rapp & Ferber, 1993) die bedingte Wahrscheinlichkeit noch durch die Wahrscheinlichkeit $p(j)$ hoch α dividiert.

$$r_{i,j} = \frac{p(j|i)}{p(j)^\alpha} \quad (3.5)$$

Als assoziative Antwort auf ein Stimuluswort i wurde das Wort j vorhergesagt, für das der Term $r_{i,j}$ in Gleichung 3.5 maximal wurde. Bei einem α von 0.68 ergaben sich die besten Voraussagen (Wettler, Rapp & Ferber, 1993), (Rapp, 1996).

Nach der Berechnung der Assoziationsstärken erfolgte ein Vergleich zwischen den berechneten Primärassoziationen und den Primärassoziationen der Teilnehmer aus der Erhebung von Russell (Russell, 1970). Bei 19 Stimuluswörtern stimmten die berechnete Primärantwort und die Primärantwort aus der Erhebung von Russell & Meseck überein. In der Erhebung produzierten die Teilnehmer im Mittel 22,5 Primärantworten. Dies bedeutet, dass nur ein geringer Unterschied zwischen den berechneten und den beobachteten Primärantworten bestand.

In (Wettler, Rapp & Sedlmeier, 2005) wurden englische Assoziationsnormen verwendet, die im Edinburgh Associative Thesaurus zur Verfügung gestellt wurden. Für die Erhebung von diesen Assoziationsnormen wurden 8210 Stimuluswörter verwendet. In (Wettler, Rapp & Sedlmeier, 2005) wurden von diesen 8210 Stimuluswörtern nur die 100 Stimuluswörter verwendet, die ursprünglich von Kent und Rosanoff zusammengestellt worden waren und die auch in der Erhebung von Russell verwendet wurden. Dies waren dieselben Stimuluswörter, die schon in (Wettler, Rapp & Ferber, 1993) verwendet wurden, nur das diesmal nicht die deutschen sondern die englischen Wörter verwendet wurden. Für die Ermittlung der Kookkurrenzen diente das British National Korpus (BNC) als Textkorpus. Es besteht aus ca. 100 Millionen Wörtern. Für die Vorhersage der Wortassoziationen wurde das gesamte Textkorpus eingelesen. Wenn eines der 100 Stimuluswörter eingelesen wurde, so wurden alle Wörter innerhalb einer Distanz von 20 Wörtern als Kookkurrenzen gezählt. Diese Distanz entspricht der durchschnittlichen Satzlänge im

Textkorpus.

Der Worthäufigkeitseffekt wurde diesmal nicht mittels Gleichung 3.5, wie in (Wettler, Rapp & Ferber, 1993) angegeben, korrigiert. Für die Korrektur des Worthäufigkeitseffekts wurde zuerst untersucht, welche Worthäufigkeiten die assoziativen Antworten der Teilnehmer im Textkorpus hatten. Dabei stellte sich heraus, dass die Teilnehmer Wörter mit mittleren Worthäufigkeiten als assoziative Antwort bevorzugten (Wettler, Rapp & Sedlmeier, 2005). Für die Korrektur des Worthäufigkeitseffekts wurden alle Wörter in Wortklassen eingeteilt, indem der Zweierlogarithmus ihrer Häufigkeiten im BNC ermittelt wurde. So entstand für jede 2er Potenz eine Wortklasse. In der Wortklasse 1 standen die Wörter mit Häufigkeit 1, in der Wortklasse 2 alle Wörter mit Häufigkeit 2, in der Wortklasse 3 alle Wörter mit den Häufigkeiten 3 oder 4, in der Wortklasse 4 alle Wörter mit den Worthäufigkeiten 5-8 usw. Nach dieser Aufteilung der Wörter in Wortklassen nach Worthäufigkeit wurde für jedes Wort einer Wortklasse ermittelt, wie häufig es von den Teilnehmern als assoziative Antwort im Assoziationsversuch gegeben wurde. Diese Häufigkeit wurde anschließend durch die Anzahl der Wörter in der zugehörigen Wortklasse dividiert. Je größer dieser Quotient ist, um so größer ist die Wahrscheinlichkeit, dass ein Wort mit dieser Häufigkeit als Responsewort im Assoziationsversuch genannt worden ist. Die Gleichung 3.4 wurde mit diesem Quotient multipliziert, um den Worthäufigkeitseffekt zu korrigieren.

Wortassoziationen können in zwei Klassen aufgeteilt werden, in syntagmatische und in paradigmatische Assoziationen. Bei den paradigmatischen Assoziationen treten die beiden assoziierten Wörter üblicherweise im selben Kontext auf. Das eine Wort kann durch das andere Wort ersetzt werden, ohne dass dies zu einem falschen Satz führt, z.B Mann und Frau. Bei syntagmatischen Assoziationen kann das eine Wort nicht durch das andere ersetzt werden. Das Stimuluswort und das Responsewort tritt normalerweise an direkt aufeinanderfolgenden Positionen im Satz auf, z.B grün und Wiese. Erwachsene Menschen produzieren mehr paradigmatische assoziative Antworten als Kinder. Während über 74 Prozent der Erwachsenen paradigmatische Assoziationen produzieren, produzieren Kinder ca. 72 Prozent syntagmatische Assoziationen (Brown & Berko, 1960).

Aufgrund der Definition von syntagmatischen und von paradigmatischen Beziehungen würde man erwarten, dass in den Arbeiten von (Wettler, Rapp & Ferber, 1993), (Wettler, Rapp & Sedlmeier, 2005) nur paradigmatische Assoziationen vorausgesagt werden können, da diese auf Kontiguitäten basieren. Dies war aber nicht der Fall. Das Modell konnte sowohl syntagmatische als auch paradigmatische Primärantworten voraussagen. Somit treten auch paradigmatische Wortpaare als Kookkurrenzen auf. Für 29 der 100 Stimuluswörter wurde die Primärantwort der Teilnehmer vom Modell richtig

vorausgesagt (Wettler, Rapp & Sedlmeier, 2005). Im Vergleich dazu produzierte ein durchschnittlicher Teilnehmer 28 richtige Primärantworten zu den 100 Stimuluswörtern. 64 von den 100 vorhergesagten Primärantworten wurden von mindestens einem Teilnehmer genannt. Im Vergleich dazu produzierte ein Teilnehmer im Mittel 72 Antworten, die mindestens noch von einem anderen Teilnehmer genannt wurden.

Ferber, Wettler & Rapp (Ferber, Wettler & Rapp, 1995) konnten ihr auf Kookkurrenzen basierendes assoziatives Wortnetz erfolgreich im Bereich des Information Retrieval einsetzen. Hierbei sollte die Suchfragenformulierung von professionellen Datenbankrechercheuren simuliert werden. Das Material bestand aus 40 schriftlichen Anfragen von Nutzern wissenschaftlicher Literaturdatenbanken. Diese Anfragen wurden einmal von professionellen Datenbankrechercheuren bearbeitet und einmal von einem Computerprogramm. Die in den Anfragen verwendeten Begriffe wurden zunächst im assoziativen Wortnetz aktiviert. Nach einem *spreading activation* Verfahren wurden die Aktivitäten auf alle Knoten (Begriffe) im assoziativen Netz verteilt. Ferber, Wettler & Rapp (Ferber, Wettler & Rapp, 1995) konnten zeigen, dass die Begriffe am stärksten aktiviert wurden, die von den Datenbankrechercheuren als Suchbegriffe verwendet wurden.

Fazit: Wettler, Rapp & Ferber (1993) können mit ihrem assoziativen Wortnetz die Voraussage von menschlichen Wortassoziationen erklären. Sie rechtfertigen ihren Ansatz mit einer assoziativen Lerntheorie, bei der die zu erwartende Assoziationsstärke gleich der bedingten Wahrscheinlichkeit ist, dass bei gegebenem Wort i das Wort j auftritt (siehe Gleichung 3.3). Die zu erwartende Assoziationsstärke lässt sich aber nur aus den Gleichungen 3.1 und 3.2 herleiten, wenn stochastische Unabhängigkeit vorausgesetzt wird. Wörter in der natürlichen Sprache sind aber nicht stochastisch unabhängig voneinander. Dies kann man z.B. am Auftreten von Substantiven und Adjektiven sehen. In unmittelbarer Nähe von einem Adjektiv ist mit sehr hoher Wahrscheinlichkeit auch ein Substantiv zu finden (z.B. großes Haus), während ein Substantiv auch ohne Adjektiv auftreten kann.

Die Verwendung von Gleichung 3.3 führte zu einem starken Worthäufigkeitseffekt, d.h. Funktionswörter wurden häufig als assoziative Antworten berechnet, so dass die Gleichung 3.3 noch mit einem Faktor gewichtet werden musste, damit die berechneten Assoziationen nicht zu stark durch den Worthäufigkeitseffekt verfälscht wurden.

3.4 HAL Hyperspace Analogue to Language - Burgess

In dem HAL Modell (Burgess, 1998) werden Ähnlichkeiten zwischen Wörtern durch einen metrischen Raum beschrieben. Dieser Raum wird auf der Grundlage des gemeinsamen Auftretens von Wörtern berechnet. Als Ergebnis erhält man einen Kontextraum, in dem die Wörter, die ähnlich sind, eine geringe Distanz zueinander haben.

Für die Bestimmung der Kookkurrenzen im HAL Modell wurden Texte aus dem UseNet (Newsgroups) als Textkorpus verwendet (Burgess, 1998). Das Textkorpus enthielt ca. 300 Millionen Wörter. Da im UseNet normale Alltagssprache verwendet wird, eignet es sich nach Burgess (Burgess, 1998) besonders als Textgrundlage für die Bestimmung von Wortähnlichkeiten.

Für die Erstellung von HAL wurden 70000 häufige Wörter bzw. Pseudowörter aus dem Textkorpus extrahiert. Diese Symbole stellten eine Art Vokabular dar. Für alle Wortpaare aus dem Vokabular wurden Kookkurrenzen mit Hilfe der Fenstertechnik, die im obigen Abschnitt Fenstertechnik beschrieben wurde, gezählt. Zur Wortpaarbildung wurde die 2. Variante, die im Abschnitt Fenstertechnik beschrieben wurde, verwendet. Das Fenster bestand aus insgesamt 21 Wörtern, wobei das Wort in der Mitte das ausgewählte Wort war. Somit wurden zehn Wörter vor dem ausgewählten Wort und zehn Wörter nach dem ausgewählten Wort bei der Wortpaarbildung berücksichtigt. Die Kookkurrenzen von zwei Wörtern wurden nicht nur gezählt, sondern mit einem Faktor gewichtet. Diese Gewichtung erfolgte umgekehrt proportional zum Abstand der Wörter zueinander. Durch diese Gewichtung erhielten Wörter mit einem geringeren Abstand zum Wort in der Mitte des Fensters einen höheren Wert als Wörter, die einen größeren Abstand zum Wort in der Mitte des Fensters hatten. Direkte Nachbarn erhielten den höchsten Wert, der der Fensterlänge entsprach. Kam ein Wort im Fenster mehrmals vor, so addierten sich die Werte. Als Ergebnis erhielt Burgess eine *Wort x Wort* Matrix, die in den Spalten die Anzahl der Kookkurrenzen im Fenster vor dem Wort enthält. Entsprechend stehen in den Zeilen die Anzahl der Kookkurrenzen im Fenster nach dem Wort. Die *Wort x Wort* Matrix enthielt insgesamt 70000 x 70000 Kookkurrenzwerte. Dies waren die Ausgangsdaten für die Simulation von Experimenten oder für die Extraktion von semantischen Wissen.

Ein Kookkurrenzvektor im HAL Modell besteht aus der Spalte und der Zeile des zugehörigen Wortes in der *Wort x Wort* Matrix. Spaltenvektor und Zeilenvektor werden einfach hintereinander gehängt und ergeben so den ganzen Kookkurrenzvektor. Somit enthält ein Kookkurrenzvektor 140000 Ele-

mente, wenn 70000 Wörter verwendet werden. Die einzelnen Vektoren repräsentieren die Koordinaten im Kontextrraum. Dies bedeutet, dass jedes Wort oder Symbol aus dem Vokabular ein Punkt in diesem Raum darstellt. Die Wortähnlichkeiten werden im HAL Modell mit Hilfe der euklidischen Metrik bestimmt. Zwei Wortvektoren repräsentieren zwei Punkte im Raum. Der Abstand dieser Punkte zueinander lässt sich mit der euklidischen Metrik bestimmen. Je kleiner der Abstand zweier Worte zueinander ist, je ähnlicher sind die beiden Worte.

Für einige Simulationen wurde der Kontextrraum von 140000 Dimensionen auf 10 bis 200 Dimensionen reduziert. Dazu wurde ermittelt, welche Vektoren die meisten Informationen enthielten. Dies wurde durch die Berechnung der Zeilenvarianz und der Spaltenvarianz bestimmt. Die Vektoren mit der kleinsten Varianz enthielten die wenigsten Informationen und wurden deshalb nicht verwendet.

Wenn die Kookkurrenzvektoren visualisiert werden sollten, wurde eine Dimensionsreduktion auf zwei Dimensionen durchgeführt. Für die Dimensionsreduktion wurde eine MDS (Multidimensionale Skalierung) Algorithmus verwendet. Bei dieser Dimensionsreduktion sollen die Orginaldistanzen möglichst erhalten bleiben.

Bei vielen Simulationen wurde keine Dimensionsreduktion durchgeführt, da die Berechnung der Varianz der Zeilen und Spalten einer großen Matrix sehr aufwendig ist.

Das HAL Modell wurde von Burgess für die Kategorisierung von Objekten verwendet (Burgess, 1998). Burgess führte hierfür ein Experiment mit drei unterschiedlichen Kategorien, und zwar Tierarten, Körperteile und geographische Orte, durch (Burgess, 1998). Für jedes Wort, das zu einer der drei Kategorien gehörte, wurde ein Vektor der Länge 200 aus dem HAL Modell extrahiert. Durch eine Dimensionsreduktion auf zwei Dimensionen mittels MDS konnten die Wörter graphisch in die drei Gruppen eingeteilt werden. Die geographischen Orte befanden sich an einer anderen Stelle als die Worte, die zu den anderen beiden Kategorien gehörten. Die beiden Kategorien Tierarten und Körperteile überlagerten sich bei dem Wort *tooth*. Dies lag daran, dass *tooth* ein Körperteil bezeichnet, das bei Tieren eine große Bedeutung hat. Das gleiche Verfahren konnte auch erfolgreich auf die Extraktion von grammatikalischen Informationen angewendet werden. Als Ergebnis erhielt man zwei separate Bereiche für Verben und Substantive. Somit enthielt das HAL Modell nicht nur semantische sondern auch grammatikalische Informationen.

Lund et al. konnten zeigen, dass die Distanzen zwischen den Wörtern mit den Reaktionszeiten aus Priming-Experimenten korrelieren (Lund & Burgess, 1996).

Burgess untersuchte, ob eine Beziehung zwischen Wortassoziationen und seinen berechneten lexikalischen Kookkurrenzen bestand (Burgess, 1998). Für diese Untersuchung wurde dasselbe Korpus verwendet, das auch schon für den Aufbau von HAL verwendet wurde. Burgess verwendete 389 stark assoziierte Wortpaare aus den Palermo und Jenkins (Palermo & Jenkins, 1964) Normen, um den Spence und Owens Effekt zu replizieren (Burgess, 1998). Die Korrelation fiel jedoch etwas schwächer als bei Spence und Owens aus. Als Wortpaare wurden nur die ersten 5 Responsewörter zu einem gegebenen Stimuluswort verwendet. Dies gab Burgess als möglichen Grund dafür an, dass seine Korrelation schwächer ausfiel als die bei Spence und Owens.

In einer weiteren Untersuchung sollte der Zusammenhang zwischen den Wortähnlichkeiten und den Auftretenshäufigkeiten der stark assoziierten Wortpaare ermittelt werden. Für diese Untersuchung wurden die stark assoziierten Wortpaare in zwei Gruppen aufgeteilt, in semantische Nachbarn und Nichtnachbarn. Als semantische Nachbarn wurden von Burgess die assoziierten Wortpaare bezeichnet, die innerhalb eines Radius von 50 Wörtern im von HAL erzeugten semantischen Raum auftraten. Entsprechend waren die Nichtnachbarn die assoziierten Wortpaare, die außerhalb eines Radius von 50 Wörtern lagen. Für jedes Wortpaar wurde die Anzahl der Teilnehmer, die das zweite Wort aus dem Wortpaar als assoziative Antwort zum ersten Wort genannt hatten, mit der Auftretenshäufigkeit im Textkorpus korreliert. Aufgrund der zuvor ermittelten Korrelation zwischen Wortassoziationen und Häufigkeit des gemeinsamen Auftretens würde man erwarten, dass die Korrelation bei Nachbarn und Nichtnachbarn gleich gut war. Dies war aber nicht der Fall, die Korrelation bei Nachbarn lag bei 0.48, und die Korrelation bei Nichtnachbarn lag nur bei 0.05. Nur bei ähnlichen Wortpaaren scheint ein Zusammenhang zwischen gemeinsamer Auftretenshäufigkeit und Wortassoziation zu bestehen. Deswegen werden nach Burgess Wortassoziationen nicht besonders gut durch lokale Kookkurrenzen repräsentiert (Burgess, 1998).

Fazit: Nach Burgess korreliert die Assoziationsstärke mit dem gemeinsamen Auftreten von Wortpaaren nur für einen Teil der Wortassoziationen. Kookkurrenzen und Wortassoziationen korrelieren nur dann, wenn die Wortpaare auch semantisch ähnlich sind. Diese Schlussfolgerung von Burgess widerspricht aber den Ergebnissen, die Wettler & Rapp erzielt haben (Wettler, 2002). Sie konnten freie Wortassoziationen mit Hilfe von Kookkurrenzen voraussagen.

3.5 LSA Latent Semantic Analysis - Landauer

Bei der LSA wird angenommen, dass in einem Textkorpus semantische Strukturen verborgen (latent) sind. LSA ist sowohl eine Theorie als auch eine Methode, aus großen Textkorpora Bedeutungen von Wörtern zu extrahieren und diese in einem semantischen Raum darzustellen. LSA verwendet im Gegensatz zu den vorherigen Modellen nicht die Fenstertechnik, um den gemeinsamen Kontext zu spezifizieren. Der gemeinsame Kontext wird bei LSA durch ein Dokument oder durch einen Abschnitt festgelegt. Als Ergebnis erhält man eine *Wort \times Dokument* Matrix. Für jedes Dokument enthält diese Matrix einen Vektor. Der Vektor enthält die Wörter, die in dem Dokument enthalten sind, in normierter Form. Somit werden bei der LSA Informationen über Distributionen von Wörtern in Dokumenten auf Vektoren abgebildet, die semantische Vektoren genannt werden. Diese semantischen Vektoren bilden zusammen einen semantischen Raum. In diesem Vektorraum geben die Positionen der semantischen Vektoren Aufschluss über den Grad ihrer Ähnlichkeit. Wörter mit ähnlichen Distributionen im Textkorpus werden auch durch ähnliche Vektoren dargestellt. In welchem Kontext ein Wort gebraucht wird, bestimmt somit seine Position im semantischen Raum.

LSA wurde ursprünglich für die Verbesserung der Ergebnisse von Suchanfragen im Internet entwickelt und wird hier als LSI (Latent Semantic Indexing) bezeichnet (Deerwester, Dumais, Furnas, Landauer & Harshmani, 1990). Beim Information Retrieval (IR) gibt es Probleme bei Polysemie, d.h. ein Wort bezeichnet verschiedene Objekte z.B. Kapelle, und bei Synonymie, d.h. für ein Objekt gibt es verschiedene Worte, z.B. Auto und Wagen, in der Suchanfrage. Wenn nach einem Objekt, das durch verschiedene Worte bezeichnet werden kann, gesucht wird, so wird in der Suchanfrage nur ein Wort für das Objekt verwendet, z.B. Auto. Es sollen aber bei dem Objekt Auto in der Suchanfrage nicht nur Dokumente gefunden werden, die Auto enthalten, sondern auch die, die Wagen enthalten.

Landauer wollte mit Hilfe von LSA die Frage beantworten: Woher kennt der Mensch die Bedeutung von neuen Wörtern (Landauer & Dumais, 1997)? Die Antwort auf diese Frage lautete: Der Mensch ermittelt die Bedeutung eines neuen Wortes aus dem Kontext, in dem es steht.

Um die Bedeutung von Wörtern zu analysieren, wird bei LSA zuerst eine *Wort \times Kontext* oder *Wort \times Dokument* Matrix aufgebaut. Die Zeilen dieser Matrix enthalten die Wörter und die Spalten die Kontexte. Die Zelle (i, j) der *Wort \times Kontext* Matrix enthält die Häufigkeit des Wortes i in dem Kontext j . Bei LSA werden nicht die reinen Worthäufigkeiten verwendet,

sondern eine gewichtete Worthäufigkeit. Für die Gewichtung wird die reine Worthäufigkeit (+1) im ersten Schritt logarithmiert. Diese Logarithmierung approximiert nach Landauer (Landauer & Dumais, 1997) die Wachstumsfunktion, die beim einfachen Lernen entsteht. Für jeden Kontext wird die Wachstumsfunktion oder Lernfunktion wieder neu gestartet. Dadurch erhält man laut Landauer (Landauer & Dumais, 1997) eine Art *Spacing Effekt*. Eine Assoziation zwischen A und B ist größer, wenn A und B in zwei unterschiedlichen Kontexten zusammen vorkommt, als wenn A und B zweimal im gleichen Kontext zusammen vorkommt.

Im zweiten Schritt wird für jedes Wort die Entropie bestimmt. Die Entropie wird über alle Worthäufigkeiten eines Wortes in den jeweiligen Dokumenten gebildet. Für jede Zeile in der *Wort x Kontext* Matrix wird die folgende Summe gebildet:

$$Entropie(w) = - \sum_k p_k(w) \log p_k(w) \quad (3.6)$$

Die Zahl k gibt die Anzahl der Kontexte an. Die Wahrscheinlichkeit $p_k(w)$ berechnet sich aus der Häufigkeit von Wort w im Kontext k dividiert durch die Gesamthäufigkeit von w , d.h. von der Häufigkeit von w in allen Dokumenten. Nach der Berechnung der Entropie eines Wortes w wird jede Zeile in der *Wort x Kontext* Matrix durch den Entropiewert der zugehörigen Zeile bzw. des Wortes dividiert. Der so berechnete Wert ist ein Maß für die Assoziation zwischen Wort und Kontext. Laut Landauer leistet der obige Schritt das gleiche wie die Konditionierungsregeln, die Rescorla & Wagner (Wagner & Rescorla, 1972) beschreiben. Das Rescorla & Wagner Modell wird im Kapitel Lerntheorien beschrieben. Das Berechnen der *Wort x Kontext* Matrix entspricht laut Landauer grob einem Konditionierungsprozess. Allerdings spezifiziert er dieses nicht weiter.

Nach der Aufstellung der *Wort x Kontext* Matrix folgt eine Singulärwertzerlegung mit anschließender Dimensionsreduktion. Diese Dimensionsreduktion führt dazu, dass die assoziativen Daten aus der *Wort x Kontext* Matrix in eine verdichtete Repräsentation transformiert werden. Bei der Singulärwertzerlegung wird eine rechteckige ($m \times n$) - Matrix in drei Matrizen zerlegt (Berry, Dumais & O'Brien, 1995), (Lay, 1994).

$$A = UDV^T \quad (3.7)$$

Die Matrix U ist eine ($m \times m$) - Matrix, deren Spalten die Eigenvektoren der Matrix AA^T enthalten. Die Matrix V ist eine ($n \times n$) - Matrix, deren Spalten die Eigenvektoren der Matrix $A^T A$ enthalten. Die Matrizen U und V sind orthonormal, somit gilt folgende Beziehung:

$$UU^T = U^T U = I \quad (3.8)$$

$$VV^T = V^T V = I \quad (3.9)$$

Die Matrix D ist eine $(m \times n)$ Diagonalmatrix, d.h. nur die Diagonale enthält Werte ungleich Null. In der Diagonale der Matrix D stehen die Singulärwerte der Matrix A . Die Singulärwerte sind die Wurzeln aus den Eigenwerten der Matrizen AA^T und $A^T A$. Die Matrix I ist eine Diagonalmatrix, die nur aus Einsen in der Diagonale besteht. Die Spalten von U werden auch als linke Singulärvektoren und die Spalten von V als rechte Singulärvektoren bezeichnet. Die Singulärwerte σ_i können in der Diagonale absteigend angeordnet werden, so dass gilt: $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$. Die Zahl r gibt dabei den Rang der Matrix A an.

$$D = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 & \dots & 0 \\ \vdots & & & & & \\ \vdots & & & & & \\ 0 & 0 & \dots & \sigma_r & \dots & 0 \\ \vdots & & & & & \\ \vdots & & & & & \\ 0 & 0 & \dots & 0 & \dots & 0 \end{pmatrix} \quad (3.10)$$

Hat die Matrix r Singulärwerte, so enthalten die Matrizen U und V r - Spalten, die sich vom Null - Vektor unterscheiden. Werden die Spalten von U mit u_i und die Spalten von V mit v_i bezeichnet, so lässt sich die obige Zerlegung von A folgendermaßen beschreiben: (dyadische Zerlegung)

$$A = \sum_{i=1}^r u_i \sigma_i v_i^T \quad (3.11)$$

Man erhält eine Approximation der Matrix A , wenn die kleinsten Singulärwerte auf Null gesetzt werden (Dimensionsreduktion). Werden nur die ersten k Singulärwerte statt der r Singulärwerte von A verwendet, so entsteht eine Matrix A_k mit Rang k , die A von allen Matrizen mit dem Rang k am besten approximiert (Berry, Dumais & O'Brien, 1995). Die Matrix A_k ist die best mögliche Approximation der Originalmatrix, bei der der quadratische Fehler am kleinsten ist (least squares best fit).

$$A_k = \sum_{i=1}^k u_i \sigma_i v_i^T \quad (3.12)$$

$$A_k = U_k D_k V_k^T \quad (3.13)$$

Die Matrix U_k enthält die ersten k - Spalten der Matrix U , die Anzahl der Zeilen bleibt erhalten, U_k ist somit eine $(m \times k)$ - Matrix. Die Matrix D_k enthält nur die ersten k - Singulärwerte in ihrer Diagonale, D_k ist eine $(k \times k)$ - Matrix. Die Matrix V_k enthält die ersten k - Spalten der Matrix V , die Anzahl der Zeilen bleibt, wie bei der Matrix U_k , erhalten, somit hat V_k k - Spalten und n - Zeilen.

Die Singulärwertzerlegung wird zur Lösung von überbestimmten und unterbestimmten linearen Gleichungssystemen verwendet. Solche Gleichungssysteme haben keine eindeutige Lösung, da ihre Koeffizientenmatrix keine Inverse besitzt, und sie somit nicht nach dem gesuchten Vektor x aufgelöst werden können. Deshalb wird eine Näherungslösung der kleinsten Fehlerquadrate (Least Squares) gesucht. Diese Näherungslösung liefert die Singulärwertzerlegung, da alle drei Matrizen der Zerlegung einzeln invertierbar sind.

Bei LSA wird ein Raum mit möglichst wenigen Dimensionen gesucht, der die Texte mit wenig Informationsverlust repräsentiert, aber Rauschen entfernt. Dieser Raum entsteht durch die Singulärwertzerlegung. Das Auffinden dieses Raumes mit weniger Dimensionen entspricht dabei der Methode der linearen Regression im zweidimensionalen Fall, bei der eine Gerade durch die veräuschten Daten gesucht wird. Bei der linearen Regression werden zwei Dimensionen auf eine Dimension (Gerade) abgebildet (reduziert). Es wird die Gerade als Approximation gewählt, die im Mittel von allen Punkten den kleinsten Abstand hat.

Bei LSA wird die Anzahl der Dimensionen, auf die der ursprüngliche Raum reduziert wird, experimentell ermittelt. Die Ähnlichkeitsberechnung erfolgt bei LSA nur mit der Matrix, bei der die Dimensionen reduziert sind (Matrix A_k). Durch die Dimensionsreduktion können Wörter mit größerer oder kleinerer Häufigkeit als in der Originalmatrix auftreten. Es können aber auch Wörter, die vorher in einem bestimmten Kontext nicht auftraten, nach dem Reduktionsschritt in einem Kontext auftreten (Landauer, Foltz & Laham, 1998).

Landauer (Landauer, 2002) erklärt den Effekt des Streichens von Dimensionen, die kleine Singulärwerte haben, damit, dass Wörter in unüblichen Zusammenhängen ignoriert werden sollen. Als Beispiel nennt er einen Lernenden, der die Sprache verstehen will. Dieser soll statistisch unbedeutende Bedeutungen ignorieren. Er soll sich mehr auf den gemeinsamen Kern in den verteilten Kontexten konzentrieren. Dimensionsreduktion ist laut Landauer ein induktiver Prozess, der Dinge ähnlicher macht.

Die Ähnlichkeit zweier Vektoren wird im reduzierten Raum üblicherweise durch den Kosinus zwischen den beiden Vektoren bestimmt. Weitere verwendete Ähnlichkeitsmaße sind das Skalarprodukt und die euklidische Distanz zweier Vektoren.

LSA wurde an dem Synonym Test TOEFL (Test of English as a Foreign Language) überprüft (Landauer & Dumais, 1997). In diesem Test erhalten die Teilnehmer ein englisches Wort in seiner Grundform. Zu diesem Wort werden vier alternative Wörter angegeben. Die Teilnehmer sollen bestimmen, welches von den vier alternativen Wörtern, dem Wort in der Grundform am ähnlichsten ist.

Zur Simulation dieses Tests wurde ein Textkorporus aus der elektronischen Version von Grolier's Academic American Encyclopedia zusammengestellt. Die Encyclopedia besteht aus 30.473 Artikeln. Von jedem Artikel wurde entweder der gesamte Artikel oder nur die ersten 2000 Zeichen verwendet, wenn der Artikel mehr als 2000 Zeichen hatte. So entstand ein Textkorporus mit 4.6 Millionen Wörtern, auf dem die LSA angewendet wurde. Die resultierende *Wort x Dokument* Matrix hatte 30.473 Spalten, eine Spalte für jeden Artikel in der Enzyklopädie, und 60.768 Zeilen. In den Zeilen standen die Worttypen, die in mindestens zwei Artikeln vorkamen. Die Matrix wurde mittels SVD auf 300 Dimensionen reduziert. Für die Reduktion wurden die 300 größten Singulärwerte verwendet.

Für die Simulation des Synonym Tests TOEFL wurden 80 Elemente aus diesem verwendet. Das LSA Modell traf seine Entscheidung zwischen den vier alternativen Wörtern, indem es den Kosinus zwischen dem Wortvektor des gegebenen Wortes und jedem Wortvektor des alternativen Wortes berechnet hatte. Es wurde das Wort, von den vier alternativen Wörtern, gewählt, das den größten Kosinus - Wert zum gegebenen Wort hatte. Das LSA Modell konnte in 52.5 Prozent der Fälle, die Wörter richtig zuordnen. Im Vergleich hierzu konnte eine große Stichprobe von nicht Englisch sprechenden U.S. College Studenten im Mittel 51.1 Prozent der Wörter richtig zuordnen.

Die Simulation des Synonym Tests wurde auch dazu verwendet, die optimale Dimension für die Dimensionsreduktion zu bestimmen. Dazu wurde die Simulation des Synonym Tests mit unterschiedlichen Dimensionen durchgeführt. Die Anzahl der Dimensionen war somit ein freier Parameter, der experimentell ermittelt werden musste. Wurden nur 2 oder 3 Dimensionen verwendet, so konnten nur 13 Prozent der Wörter richtig zugeordnet werden. Bei ca. 300 Dimensionen wurden die besten Ergebnisse erzielt, es wurden 45-53 Prozent der 80 Wörter richtig zugeordnet. Wurden mehr Dimensionen verwendet, so wurden die Resultate wieder schlechter. Wurde keine Dimensionsreduktion durchgeführt, so konnten nur 16 Prozent der Wörter richtig zugeordnet werden.

Das Trainingskorpus, das beim Synonym Test verwendet wurde, enthielt ungefähr die gleiche Anzahl Wörter, die im Mittel ein 9th Grade Student gelesen hat.

LSA wurde auch zur Essaybewertung verwendet (Landauer, 2002). Für diese Aufgabe wurde ein semantischer Raum aus Essays erzeugt. Ein unbewertetes Essay wurde aufgrund seiner Ähnlichkeit zu den bereits bewerteten Essays bewertet. Die gleiche Aufgabe wurde auch von Menschen durchgeführt. Die Korrelation zwischen maschineller Bewertung und menschlicher Bewertung lag bei 0.81, während die Korrelation zwischen zwei unterschiedlichen menschlichen Bewertungen bei 0.83 lag.

Griffiths und Steyvers verwendeten LSA für die Vorhersage von Wortassoziationen (Griffiths & Steyvers, 2002). Als Textkorpus wurde das TASA Korpus verwendet. Es wurden alle Dokumente des TASA Korpus (37.651) verwendet und alle Wörter, die mindestens zehn mal im Textkorpus vorkamen. Stop Wörter, dies sind häufige Funktionswörter, wurden dabei nicht verwendet. Insgesamt wurden 26.414 Wörter aus dem TASA Korpus für den Aufbau einer *Wort x Dokument* Matrix verwendet. Als Assoziationsnormen dienten die Normen von Nelson, McEvoy und Schreiber (Nelson, McEvoy & Schreiber, 1999). In den 26.414 Wörtern waren 4544 Wörter aus den Normen enthalten. Der Kosinus des Winkels zwischen zwei Wortvektoren wurde als Maß für die Assoziationsstärke von zwei Wörtern verwendet. Die Assoziationsstärken aller Wortpaare aus den Normen wurden bestimmt. Zur Bestimmung der Assoziationsstärke wurde der Kosinus zwischen jedem Wort und den anderen 4543 Wörtern aus den Normen berechnet. Für jedes Stimuluswort wurden die Rangplätze der ersten zehn assoziierten Wörter aus den Assoziationsnormen bestimmt. Das Responsewort mit dem größten Kosinus Wert zum gegebenen Stimuluswort erhielt den Rangplatz 1. Das Responsewort mit dem zweitgrößten Kosinus Wert erhielt den Rangplatz 2 usw. . Danach wurde für jedes Stimuluswort der Rangplatz der Primärantwort aus den Assoziationsnormen bestimmt. Der Median Rang für die Primärantwort betrug 32. Die besten Ergebnisse bei der Vorhersage der freien Wortassoziationen wurden bei einer Dimensionsreduktion auf 500 Dimensionen erzielt. LSA sagte 507 Primärantworten von 4544 Wörtern richtig voraus, d.h. 507 Primärantworten standen auf Rangplatz 1 (Griffiths & Steyvers, 2002).

3.6 Topic Modell - Griffiths & Steyvers

Griffiths & Steyvers hatten LSA zur Voraussage von Wortassoziationen verwendet (siehe vorherigen Abschnitt). Sie wollten die Voraussage von Wortassoziationen verbessern. Dazu entwickelten sie ihr Topic Modell. Wie bei

Landauer und Dumais waren die Ausgangsdaten eine *Wort x Dokument* Matrix.

Beim Topic Modell wird angenommen, dass ein Dokument verschiedene Themen enthält. Diese Themen wiederum enthalten bestimmte Wörter aus dem Dokument. In einem Topic Modell wird jedes Dokument als eine Wahrscheinlichkeitsverteilung von Themen repräsentiert. Jedes Thema (Topic) ist wiederum eine Wahrscheinlichkeitsverteilung von Wörtern. Topic Modelle können einmal zur Dokumentgenerierung und einmal zur Extraktion von Themen verwendet werden. Die Dokumentgenerierung ist ein probabilistischer Prozess, bei dem die Themen und die darin enthaltenen Wörter vorgegeben sind. Gesucht ist ein Dokument, das mit einer vorgegebenen Wahrscheinlichkeit die Themen enthält.

Z.B Zur Dokumentgenerierung von Dokument A stehen zwei Themen Thema 1 und Thema 2 zur Verfügung. Mit einer Wahrscheinlichkeit von 0.8 soll ein Wort aus dem Thema 1 und mit einer Wahrscheinlichkeit von 0.2 soll ein Wort aus dem Thema 2 in dem Dokument A enthalten sein. Das Dokument A besteht somit aus einer Mixtur von Wörtern aus den Themen 1 und 2.

In einem umgekehrten Prozess können aus den Dokumenten die enthaltenen Themen und die Wahrscheinlichkeit, dass ein Wort aus einem bestimmten Thema in einem gegebenen Dokument vorkommt, ermittelt werden. Dieser umgekehrte Prozess wurde von Griffiths und Steyvers für die Vorhersage von Wortassoziationen verwendet.

Bei der Vorhersage von Wortassoziationen soll zu einem gegebenen Wort bestimmt werden, welche anderen Wörter ebenfalls auftreten, wenn das gegebene Wort auftritt.

Griffiths und Steyvers nehmen für ihr Modell an, dass die unterschiedlichen Dokumente eine konstante Anzahl T von Themen enthalten (Griffiths & Steyvers, 2002). Die Verteilung der Wörter in einem Dokument ist dann eine Mixtur aus diesen T Themen. Mit einer Wahrscheinlichkeit von $P(w_i)$ kommt ein Wort w_i im Dokument vor. Diese Wahrscheinlichkeit wird folgendermaßen bestimmt:

$$P(w_i) = \sum_{j=1}^T P(w_i|z_i = j)P(z_i = j) \quad (3.14)$$

Die Variable z_i gibt an, aus welchem Thema das i -te Wort ausgewählt wird. $P(w_i|z_i = j)$ ist die Wahrscheinlichkeit, dass das i -te Wort aus dem j -ten Thema ausgewählt wird. $P(z_i = j)$ gibt an, mit welcher Wahrscheinlichkeit das Thema j in einem Dokument enthalten ist. Der Inhalt der Themen wird in $P(w|z = j)$ wiedergegeben, $P(w|z = j)$ gibt an welche Wörter für

ein Thema wichtig sind, während $P(z)$ angibt, wie die verschiedenen Themen in einem Dokument vermischt sind. Z.B. Es existieren nur zwei Themen A und B, dann gibt $P(w|A)$ die Wahrscheinlichkeit an, dass das Wort w zum Thema A gehört. Die Wahrscheinlichkeit, dass das Wort w zum Thema B gehört, ist dann $P(w|B) = 1 - P(w|A)$, da das Wort w zu einem der beiden Themen gehören muss. Mit einer Wahrscheinlichkeit von $P(A)$ ist das Thema A im Dokument vertreten. Entsprechend ist das Thema B mit einer Wahrscheinlichkeit von $P(B) = 1 - P(A)$ im Dokument vertreten.

Jedes Wort w_i gehört zu einem Dokument d_i . $\theta^{(d_i)}$ gibt die Themenverteilung für das Dokument d_i an. Mit $\theta_j^{(d_i)}$ wird die Wahrscheinlichkeit, mit der das j -te Thema im Dokument d_i vorkommt, bezeichnet.

$$\theta_j^{(d_i)} = P(z_i = j) \quad (3.15)$$

Das j -te Thema ist eine Distribution über alle Wörter aus dem Vokabular.

$$\phi_{w_i}^{(j)} = P(w_i | z_i = j) \quad (3.16)$$

Mit der Wahrscheinlichkeit $\phi_{w_i}^{(j)}$ ist das Wort w_i im Thema j enthalten.

Griffiths und Steyvers verwendeten die Methode Markov Chain Monte Carlo mit einem symmetrischen Dirichlet (α) prior für die Bestimmung von $\theta_j^{(d_i)}$ für jedes Dokument d_i und einem symmetrischen Dirichlet (β) prior für die Bestimmung von $\phi_{w_i}^{(j)}$ für jedes Thema (Griffiths & Steyvers, 2002).

Griffiths und Steyvers verwendeten die Methode Markov Chain Monte Carlo mit Gibbs Sampling, um die Menge der Themen aus einem großen Textkorpus zu extrahieren. Mit Hilfe des Gibbs Sampling Algorithmus wurde jedem Wort im Textkorpus die Wahrscheinlichkeit, dass es zu einem bestimmten Thema gehört, zugeordnet. Als Textkorpus wurde ein Teil des TASA Textkorpus verwendet, das 5.628.867 Wörter enthielt. Wie bei LSA wurde zuerst eine *Wort x Dokument* Matrix aufgestellt, die die gewichteten Worthäufigkeiten in einem Dokument enthielt.

Im nächsten Schritt wurde beim Topic Modell für jede Spalte die Summe über die Spalte gebildet, und anschließend wurde jedes Element in der Spalte durch die zugehörige Summe dividiert. So ergab sich eine empirische Schätzung über die Verteilungen von Wörtern in jedem Dokument.

Bei LSA wird die *Wort x Dokument* Matrix durch die Singulärwertzerlegung in drei Matrizen zerlegt. Das Topic Modell kann auch als ein Verfahren für eine Matrixzerlegung interpretiert werden (Hofmann, 1999). Die *Wort x Dokument* Matrix wird beim Topic Modell in zwei Matrizen, eine Topic Matrix und eine Dokument Matrix, zerlegt (Steyvers & Griffiths, in press).

$$P = \Phi\Theta \quad (3.17)$$

Die Spalten von Φ enthalten $\phi^{(j)}$ und die Spalten von Θ enthalten $\theta^{(d)}$ (Griffiths & Steyvers, 2002). Φ ist eine Matrix, dessen Zeilen die Wörter und dessen Spalten die Themen beinhalten. Somit ist Φ eine *Wort \times Topic* Matrix. Θ ist eine Matrix, dessen Spalten aus den Themen und dessen Zeilen aus den Dokumenten bestehen.

Diese Faktorisierung führt zu einer Dimensionsreduktion, die bei der LSA durch die Singulärwertzerlegung erfolgt. Die Anzahl der Dimensionen wird im Topic Modell nicht durch die Singulärwerte, sondern durch die Anzahl der Themen festgelegt. Beim Topic Modell sind die Vektoren der Topic Matrix und der Dokument Matrix Wahrscheinlichkeitsverteilungen. Dies ist der wesentliche Unterschied zur LSA. Bei LSA können die Vektoren auch negative Werte enthalten, dies ist beim Topic Modell nicht der Fall, da es sich um Wahrscheinlichkeitsverteilungen handelt. Deswegen handelt es sich beim Topic Modell auch um eine nicht negative Matrixfaktorisierung.

Eine Möglichkeit, Wortähnlichkeiten mit dem Topic Modell zu bestimmen, besteht darin, die Ähnlichkeit der Vektoren $\theta^{(1)}$ und $\theta^{(2)}$ zu bestimmen (Steyvers & Griffiths, in press). Die Ähnlichkeit dieser beiden Vektoren kann einmal durch das Skalarprodukt oder durch den Kosinus bestimmt werden. Für die Voraussage von Wortassoziationen wurde von Griffiths und Steyvers aber ein alternativer Ansatz gewählt (Steyvers & Griffiths, in press). Die Voraussage von Wortassoziationen entspricht im Topic Modell dem Fall, dass ein Wort in einem neuen Kontext gefunden wird. Es soll jetzt vorausgesagt werden, welche anderen Wörtern in diesem neuen Kontext zu erwarten sind (Steyvers & Griffiths, in press). Dies hängt wiederum davon ab, welchem Thema das neue Wort zugeordnet wird. Wenn das Wort w_1 in einem Dokument auftritt, so kann die Auftretenswahrscheinlichkeit von einem Wort w_2 in demselben Dokument durch $P(w_2|w_1)$ berechnet werden. Für die Voraussage von Wortassoziationen wird angenommen, dass zwei assoziierte Wörter dem gleichen Thema zugeordnet werden. Diese Annahme führt zu folgenden beiden Gleichungen (Griffiths & Steyvers, 2002):

$$P_1(w_2|w_1) = \sum_z P(w_2|z)P(z|w_1) \quad (3.18)$$

$$P(z|w_1) \propto P(w_1|z)P(z) \quad (3.19)$$

Der Index in $P_1(w_2|w_1)$ gibt die Restriktion auf ein bestimmtes Thema an. Mit Hilfe von $P_1(w_2|w_1)$ sollen die Assoziationen zu einem gegebenen Wort bestimmt werden. Als Assoziationsnormen dienten die Normen von Nelson,

McEvoy und Schreiber (Nelson, McEvoy & Schreiber, 1999), die schon bei der Vorhersage von Wortassoziationen mit LSA verwendet worden waren.

Zur Vorhersage von Wortassoziationen wurden 500 Themen verwendet. Beim Topic Modell wurde für alle 4544 Wörter, die in den Assoziationsnormen enthalten waren, die Wahrscheinlichkeit $P_1(w_2|w_1)$ berechnet. Zur Rangbestimmung der assoziierten Wörter wurde das gleiche Verfahren wie bei LSA verwendet (siehe vorherigen Abschnitt). 585 Primärantworten wurden mit Hilfe des Topic Modells richtig vorausgesagt. Dies war eine deutliche Verbesserung gegenüber LSA, bei der nur 507 Primärantworten richtig vorausgesagt wurden. Der Median Rang für die Primärantwort betrug 32.

Fazit: Das Topic Modell eignete sich besser zur Voraussage von Wortassoziationen als LSA.

3.7 ICAN Modell - Lemaire & Denhiere

Lemaire und Denhiere konstruierten inkrementell aus einem maschinenlesbaren Textkorpus ein assoziatives Netzwerk (Lemaire & Deniere, 2004). Dabei verwendeten sie die Fenstertechnik, die im Abschnitt Fenstertechnik beschrieben worden ist. Die Wortpaare wurden mit Hilfe der 2. Variante aus dem Abschnitt Fenstertechnik gebildet. Das ausgewählte Wort war das Wort in der Mitte.

Zum Aufbau des assoziativen Netzwerkes wurden dieselben Gleichungen, wie beim Modell von Wettler & Rapp, das im Abschnitt Assoziatives Wortnetz von Wettler und Rapp beschrieben wurde, verwendet. Im Unterschied zu Wettler und Rapp wurde das assoziative Netzwerk von Lemaire und Denhiere inkrementell aufgebaut, während bei Wettler und Rapp die Assoziationsstärken aufgrund von Kookkurrenzen berechnet wurden.

In der Gleichung 3.1 verwendeten Lemaire und Denhiere für die Lernrate θ_1 einen Wert von 0.5. Für die Hemmungsrate θ_2 wurde ein Wert von 0.1 verwendet. Dagegen gingen Wettler und Rapp von einer gleichen Lernrate und Hemmungsrate aus, d.h. θ_1 ist gleich θ_2 .

Lemaire und Denhiere verwendeten ein Fenster, das aus 11 Wörtern bestand. Durch die Gleichung 3.1 werden die Assoziationsstärken von dem Wort in der Mitte zu allen anderen Wörtern im Fenster verstärkt. Die Gleichung 3.2 vermindert die Assoziationsstärken von dem Wort in der Mitte zu allen anderen Wörtern, die nicht im Fenster stehen. Zusätzlich zu den beiden Gleichungen 3.1 und 3.2 wurde noch eine dritte Gleichung für die Berechnung von Kookkurrenzen zweiter Ordnung zum Aufbau des assoziativen Netzwerkes verwendet. Lemaire und Denhiere definieren Kookkurrenzen zweiter Ordnung folgendermaßen: Das Wort A ist eine Kookkurrenz zweiter Ordnung

von B, wenn es mit einem Wort C gemeinsam auftritt, das zusammen mit B auftritt. Kookkurrenzen höherer Ordnung werden bei LSA mittels Singulärwertzerlegung berücksichtigt (Lemaire & Deniere, 2004). Kookkurrenzen höherer Ordnung spielen nach Lemaire und Deniere eine wichtige Rolle in der latenten Struktur der Wortbenutzung. Zwei Wörter können miteinander assoziiert sein, obwohl sie nie zusammen auftreten, z.B. Internet und Web (Lemaire & Deniere, 2004).

Die Kookkurrenzen zweiter Ordnung wurden nach folgenden Algorithmus berechnet:

Sei $a_{i,k}$ das Gewicht einer Verbindung von einem Wort i nach einem Wort k . Existiert eine Verbindung von k nach j mit dem Gewicht $a_{k,j}$, so erzeuge eine neue Verbindung von i nach j , falls diese noch nicht existiert, mit dem Gewicht $a_{i,k} \cdot a_{k,j}$.

Eine Verbindung $i \rightarrow k \rightarrow j$ führt somit zu einer neuen Verbindung $i \rightarrow j$.

Existiert schon eine Verbindung von i nach j , so wird diese folgendermaßen verstärkt:

$$a_{i,j} = a_{i,j} + \theta_3 \cdot (1 - a_{i,k}) \cdot (a_{i,k} \cdot a_{k,j}) \quad (3.20)$$

Die Variable θ_3 ist ein freier Parameter. Durch die Regel für Kookkurrenzen zweiter Ordnung werden die Nachbarn vom Wort k ebenfalls mit dem Wort i verbunden, falls sie dies noch nicht sind. Sind die Nachbarn vom Wort k bereits mit i verbunden, so wird deren Verbindung verstärkt. Dies wird von Lemaire und Deniere der Kookkurrenz Effekt zweiter Ordnung genannt.

Wenn die Gewichte oder Assoziationsstärken zweier Wörter einen gewissen Schwellwert unterschreiten (durch die Anwendung von Gleichung 3.2), so wird die Verbindung wieder gelöscht, d.h. diese Wörter erhalten die Assoziationsstärke Null.

Die Ähnlichkeiten zwischen Wörtern werden in dem Modell von Lemaire und Deniere nicht symmetrisch repräsentiert. Ihr Netzwerk besteht aus Wörtern und orientierten numerischen Verbindungen zwischen den Wörtern. Die numerischen Verbindungen zwischen den Wörtern repräsentieren die semantische Ähnlichkeit oder Assoziationsstärke von Wörtern.

Die Ähnlichkeiten von zwei Wörtern W_1 und W_2 werden im ICAN Modell folgendermaßen bestimmt:

- Existiert eine direkte Verbindung von dem Wort W_1 nach W_2 , so wird das Gewicht der Verbindung als Ähnlichkeitsmaß verwendet.
- Existiert keine direkte Verbindung von dem Wort W_1 nach W_2 , aber eine indirekte Verbindung über ein weiteres Wort Z , d.h. $W_1 \rightarrow Z \rightarrow W_2$,

so wird als Ähnlichkeitsmaß eine Kombination von den zwei Gewichten der beiden Verbindungen gewählt, z.B. das Produkt der beiden Gewichte.

- Existiert weder eine direkte Verbindung von dem W_1 nach W_2 , noch eine indirekte Verbindung über ein weiteres Wort Z , so sind die beiden Wörter W_1 und W_2 nicht ähnlich, sie bekommen das Ähnlichkeitsmaß Null.

Dadurch, dass die Gewichte der Verbindungen W_1 nach W_2 und W_2 nach W_1 unterschiedlich sein können, wird die asymmetrischen Eigenschaft der menschlichen Entscheidung von Ähnlichkeit Rechnung getragen (Lemaire & Deniere, 2004).

Lemaire und Deniere hatten ihr Modell ICAN anhand von französischen Assoziationsnormen überprüft. Sie verwendeten einen französischen Textkorpus, das ca. 3.2 Millionen Wörtern enthielt. Das Textkorpus bestand aus Texten, die für Kinder geschrieben waren. Dies waren Geschichten, Fabeln, Enzyklopädien und Textbücher für Kinder. Für die Berechnung der Wortähnlichkeiten wurden nur Wörter verwendet, die mindestens dreimal im Textkorpus vorkamen. Funktionswörter wurden nicht verwendet. Als Assoziationsnormen wurden die Normen, die von de la Haye erhoben wurden, verwendet. Die Normen wurden mit 9 jährigen Kindern erhoben (Lemaire & Deniere, 2004).

Nachdem das assoziative Netzwerk mit Hilfe von ICAN gebildet worden war, wurde die Ähnlichkeit zwischen 200 Wörtern und den 6 zugehörigen assoziierten Wörtern, jeweils die ersten drei und die letzten drei assoziierten Wörter, aus den Assoziationsnormen berechnet. 16 Wörter waren nicht im Textkorpus enthalten, deshalb wurden nur 1184 Wortpaare verwendet. Die berechneten Ähnlichkeiten wurden mit dem Assoziationswert, der sich aus den Normen ergab, verglichen. Der Assoziationswert gibt an, wie viel Prozent der Teilnehmer die jeweilige assoziative Antwort gegeben haben.

Lemaire und Deniere hatten zwei Hypothesen, die sie testen wollten.

- Ihr Modell sollte zwischen den ersten drei assoziierten Wörtern und den letzten drei assoziierten Wörtern unterscheiden. Die berechneten Ähnlichkeitswerte sollten dabei absteigend, von dem ersten assoziierten Wort (Primärantwort) bis zum letzten assoziierten Wort, dies ist das Wort, das die wenigsten Teilnehmer genannt hatten, ausfallen.
- Es sollte eine gute Korrelation zwischen den menschlichen Daten und den Daten, die das Modell berechnet hatte, entstehen.

Die beste Korrelation von menschlichen Daten und Daten vom Modell wurde bei einer Fensterlänge von 11 erzielt, der freie Parameter θ_3 hatte dabei den Wert 0.02. Wenn die Assoziationsstärke von 0.1 unterschritten wurde, so wurde die Verbindung gelöscht, d.h. es wurde ein Schwellwert von 0.1 gewählt. Der Korrelationskoeffizient zwischen menschlichen Daten und Daten vom Modell betrug $r(1184)=0.5$, d.h. es bestand eine signifikante Korrelation zwischen menschlichen Daten und Modelldaten (Lemaire & Deniere, 2004).

Um die erste Hypothese zu bestätigen, berechneten Lemaire und Deniere den Durchschnittswert der Ähnlichkeitswerte zwischen den Stammwörtern und den zugehörigen assoziierten Wörtern. Dasselbe wurde auch mit den Daten der Teilnehmer durchgeführt. Dasselbe Textkorpus, das bei ICAN verwendet wurde, wurde auch für die Anwendung von LSA verwendet. Somit konnte ICAN mit LSA verglichen werden. Die Ergebnisse sind in der Tabelle 3.1 zusammengefasst.

	1 st associates	2 nd associates	3 rd associates	Last associates
ICAN	0.415	0.269	0.236	0.098
Norms	30.5	13.5	8.2	1
LSA	0.26	0.23	0.19	0.11

Tabelle 3.1: Vergleich: ICAN Modell, LSA Modell und Assoziationsnormen, Tabelle entnommen aus (Lemaire & Deniere, 2004).

Die Werte von ICAN waren signifikant unterschiedlich, außer bei dem zweiten und dem dritten assoziierten Wort, hier betrug der Unterschied lediglich 10 Prozent. Die erste Hypothese wurde somit von ICAN erfüllt.

Im Vergleich mit LSA lieferte ICAN stärkere Ähnlichkeitswerte zwischen dem Stimuluswort und der zugehörigen Primärantwort. Die Korrelationen von menschlichen Daten und von LSA waren ebenfalls schlechter als bei ICAN. Der Korrelationskoeffizient lag nur bei $r(1184)=0.39$.

Lemaire und Deniere testeten den Einfluss, den die Kookkurrenzen zweiter Ordnung in ihrem Modell hatten, indem sie die Regel für die Kookkurrenzen zweiter Ordnung wegliessen. Ohne diese Regel verminderte sich die Korrelation zwischen menschlichen Daten und Modelldaten von 0.5 auf 0.48. Somit hatten die Kookkurrenzen zweiter Ordnung keinen großen Einfluss auf die Simulationen.

In einer weiteren Simulation sollte untersucht werden, ob die Kookkurrenzhäufigkeit ein gutes Modell für die Wortähnlichkeit darstellt. Hierfür wurde ICON nur mit der Lernregel (siehe Gleichung 3.1), die die direkten Kook-

kurrenzen berücksichtigt, gestartet. Lemaire und Denhiere stellten hierbei eine Verschlechterung der Korrelation zwischen Modelldaten und menschlichen Daten fest ($r(1184)=0.39$). Dies lag ihrer Meinung daran, dass die Kookkurrenzhäufigkeit zu einer Überschätzung der semantischen Ähnlichkeit geführt hatte (Lemaire & Deniere, 2004).

Fazit: Lemaire und Denhiere verwendeten ein inkrementelles Lernmodell zur Voraussage von Wortassoziationen, dessen Kern die Gleichungen aus dem Estes Modell enthielt. Um die Güte der berechneten Assoziationen zu bestimmen, hatten sie einmal die Assoziationswerte, die aus den Assoziationsnormen bestimmt wurden, mit den zugehörigen Ähnlichkeitswerten von ihrem Modell korreliert. In einer weiteren Untersuchung hatten sie die Assoziationswerte von den 6 verwendeten Assoziierten aus den Normen mit den zugehörigen Ähnlichkeitswerten von ihrem Modell verglichen, indem sie für die 6 Assoziierten den Durchschnittswert vom Assoziationswert mit dem Durchschnittswert vom Ähnlichkeitswert verglichen hatten (siehe dazu Tabelle 3.1). Bei der Korrelation und bei dem Vergleich wurden jedoch nur die 1184 Wortpaare aus den Assoziationsnormen berücksichtigt. Es wurden nicht die Ähnlichkeitswerte von anderen Wortpaaren, die zwar ein Stimuluswort aus den Normen aber ein Responsewort enthielten, das nicht zu den 6 verwendeten Assoziierten aus den Normen gehörte, untersucht.

Lemaire und Denhiere machten keine Angabe dazu, wie viele Primärantworten ihr Modell richtig vorausgesagt hatte. Sie berechneten nur die Assoziationsstärke vom Stimuluswort zur Primärantwort aus den Assoziationsnormen. Sie untersuchten nicht, ob ihr Modell eine andere Primärantwort berechnet hatte.

3.8 Fazit

Alle vorgestellten Methoden konnten im gewissen Maße Wortassoziationen voraussagen. Sie verwendeten dazu unterschiedliche Auszählungsalgorithmen und Textkorpora.

Sie unterscheiden sich einmal in der Art, wie die Assoziationen berechnet wurden und in der Art, wie die Güte der Assoziationen bestimmt wurde. Assoziationen können aufgrund der Anzahl von Kookkurrenzen in Fenstern oder Dokumenten oder durch die Anwendung von inkrementellen Lernalgorithmen berechnet werden. Nach der Berechnung der Assoziationen muss getestet werden, wie gut stimmen die berechneten Assoziationen mit den menschlichen Assoziationen überein. Die Bestimmung der Güte der Assoziationen erfolgt entweder anhand von Korrelationen oder anhand der Anzahl richtig vorausgesagter Primärantworten. Für die Korrelation wird meistens die berechnete

Assoziationsstärke und die Häufigkeit, mit der das Resonsewort zum Stimuluswort in den Assoziationsnormen genannt wurde, verwendet. Bei diesem Verfahren werden aber nicht mögliche Responsewörter berücksichtigt, die in den Assoziationsnormen nicht vorkommen (Wettler, 2002). Um diese zu berücksichtigen, muss eine Korrelation über alle möglichen Wortpaare gebildet werden. Bei dem assoziativen Wortnetz von Wettler & Rapp wären dies 6.3 Millionen Wortpaare (Wettler, 2002). Dies ist aber laut Wettler nicht sinnvoll, da 90 Prozent aller Wortpaare niemals zusammen auftreten (Wettler, 2002). Deswegen verwendeten Wettler und Rapp als Maß für die Güte ihrer vorausgesagten Assoziationen die Anzahl richtig vorausgesagter Primäranworten. In dieser Arbeit soll ebenfalls dieses Gütemaß verwendet werden. Zu diesem Gütemaß werden noch weitere drei Kriterien hinzugefügt, um das Gütemaß weiter zu verfeinern. Das verwendete Gütemaß wird in dem Kapitel Simulationen im Abschnitt Kriterien beschrieben.

Die Frage ist, wie kommt es dazu, dass diese unterschiedlichen Modelle alle Wortassoziationen voraussagen können. Ich stelle die Hypothese auf, Wortassoziationen sind die Ergebnisse von Lernprozessen. Um dies zu überprüfen, werden drei bekannte Lerntheorien aus der Psychologie zur Simulation des Erlernens von Wortassoziationen verwendet. Im nächsten Kapitel werden diese drei Lerntheorien vorgestellt.

Kapitel 4

Lerntheorien

In diesem Kapitel werden die drei Lerntheorien, die zur Berechnung von Wortassoziationen verwendet werden sollen, vorgestellt. Alle drei Lerntheorien haben einen gemeinsamen Kern, das Assoziationsgesetz. Nach dem Assoziationsgesetz werden Dinge in unmittelbarer zeitlicher Nähe miteinander verknüpft, z.B. Donner und Blitz.

Die Lerntheorien unterscheiden sich unter anderem darin, dass sie für unterschiedliche Situationen entwickelt wurden. Bei den Lerntheorien von Rescorla & Wagner und von Gallistel ist die Situation das klassische Konditionierungsexperiment. Die Lerntheorien von Rescorla & Wagner und Gallistel haben das Ziel möglichst viele Phänomene, die in klassischen Konditionierungsexperimenten beobachtet wurden, mit ihrer Theorie zu erklären.

Das Estes Modell betrachtet dagegen eine allgemeine Situation. Der Mensch oder das Tier nimmt bestimmte Merkmale aus seiner Umgebung wahr und reagiert auf diese wahrgenommenen Merkmale. Nach der Reaktion tritt ein Ereignis ein. Nach dem Eintreten des Ereignisses werden die wahrgenommenen Merkmale aus der Umgebung mit der Reaktion auf diese verknüpft. Ziel dieser Lerntheorie ist es zu erklären, wie aufgrund einer aktuellen Situation vorausgesagt werden kann, ob ein bestimmtes Ereignis eintreten wird oder nicht eintreten wird.

Bei den beiden Lerntheorien von Rescorla & Wagner und Gallistel, in der die Situation ein klassisches Konditionierungsexperiment ist, werden konditionierte und unkonditionierte Stimuli betrachtet. Das klassische Konditionierungsexperiment wurde erstmals von dem Psychologen Iwan Pawlow durchgeführt und beschrieben.

4.1 Bedingter Reflex

Beim Klassischen Konditionieren wird ein unkonditionierter Stimulus (US) mit einem neutralen Stimulus gemeinsam dargeboten. Ein unkonditionierter Stimulus kann z.B. Nahrung sein, dessen Wahrnehmung zu vermehrter Speichelbildung führt. Die Reaktion auf einen unkonditionierten Stimulus wird unkonditionierte Reaktion (UR) oder unbedingte Reaktion genannt. Auf den US Futter ist die UR vermehrte Speichelsekretion. Unkonditionierte Reaktionen sind angeboren, sie werden auch als unbedingte Reflexe bezeichnet. Ein neutraler Stimulus kann z. B. eine Glocke oder ein Lichtsignal sein. Als neutrale Stimuli können beliebige Reize verwendet werden, die nicht zu einer unkonditionierten Reaktion (UR) führen.

In seinen Versuchen mit Hunden ließ Pawlow vor der Fütterung eine Glocke ertönen. Nach mehrmaliger Wiederholung dieses Versuchs, Paarung von dem US Futter mit dem neutralen Stimulus Glocke, wurde dem Versuchstier nur die Glocke als Reiz dargeboten, aber kein Futter, und die vermehrte Speichelsekretion des Hundes (UR) trat ebenfalls ein, obwohl der US Futter nicht dargeboten wurde. Durch die Paarung von dem US Futter mit dem neutralen Stimulus Glocke war aus dem neutralen Stimulus Glocke ein konditionierter Stimulus (CS) geworden. Auf den konditionierten Stimulus Glocke reagierte der Hund jetzt ebenfalls mit vermehrter Speichelsekretion. Dies ist die konditionierte Reaktion (CR) oder der bedingte Reflex auf den CS Glocke.

Der Hund hat gelernt, dass er nach dem Glockenton Futter erhält. In dem Lernvorgang wird eine Verbindung zwischen dem CS und der Reaktion auf den US aufgebaut.

Wird ein CS, wie oben beschrieben, auf einen US konditioniert, so handelt es sich um eine Konditionierung erster Ordnung. Konditionierungen höherer Ordnung bauen auf einer Konditionierung erster Ordnung auf. Eine Konditionierung zweiter Ordnung (*second order conditioning*) erfolgt in zwei Phasen: In der 1. Phase des Konditionierungsexperimentes wird ein Reiz CS_1 zusammen mit einem US dargeboten. In der 2. Phase wird ein weiterer Reiz CS_2 zusammen mit dem konditionierten Reiz CS_1 dargeboten. Der US wird in der zweiten Phase nicht dargeboten. Die konditionierte Reaktion erfolgt sowohl auf den CS_1 als auch auf den CS_2 . Der Reiz CS_2 wird als ein Reiz höherer Ordnung bezeichnet, während der CS_1 als ein Reiz erster Ordnung bezeichnet wird.

Eine konditionierte Reaktion (CR) kann aber auch wieder abgeschwächt oder ganz gelöscht werden, wenn nach der Darbietung des CS kein US mehr folgt. Die Verbindung zwischen dem CS und der Reaktion auf den US wird durch das Fehlen des US bei der Darbietung des CS abgeschwächt.

Das Erlernen einer konditionierten Reaktion und das Abschwächen oder

Löschen einer konditionierten Reaktion kann von allen drei Lerntheorien erklärt werden. In den folgenden drei Abschnitten werden die drei Lerntheorien beschrieben.

4.2 Das Estes Modell - Stimulus Sampling Theorie

Die Stimulus Sampling Theorie (SST) ist eine statistische Lerntheorie, die von W. K. Estes entwickelt wurde (Estes, 1950).

“Lernen besteht für Estes in einer Verknüpfung situativer (S-) und reaktiver (R-) Merkmale.” (K. Foppa, S. 371) (Foppa, 1965).

Situationen werden als ungeordnete Mengen aus N voneinander unabhängigen Reizelementen beschrieben. Diese stellen die gesamte Umwelt dar. Zu jedem Zeitpunkt t wird eine Teilmenge s dieser N Merkmale zufällig ausgewählt, d.h. es wird immer nur ein Teil aus der gesamten Umwelt wahrgenommen. Dies kann z.B der Versuchsraum, der Versuchsleiter, ein Geruch usw. sein. Man kann sich die Merkmale als N Kugeln in einer Urne vorstellen. In jedem Lerndurchgang wird eine bestimmte Anzahl von Kugeln aus der Urne gezogen. Dies sind die in der aktuellen Situation wahrgenommenen Merkmale. Die Ziehung der s Kugeln (Stichprobe) kann auf unterschiedliche Weise gedeutet werden:

1. Jedes Merkmal geht mit der Wahrscheinlichkeit θ in die Stichprobe ein. Dies ist unabhängig davon, wie viele Merkmale sonst noch zur Stichprobe gehören. Dies führt zu einer Binomialverteilung über die Stichprobengröße. Somit variiert die Anzahl der Merkmale in der Stichprobe bei jeder Ziehung. Die durchschnittliche Stichprobengröße beträgt $N \cdot \theta$.
2. Aus der Urne wird eine feste Zahl von s Kugeln ohne Zurücklegen gezogen. Die Wahrscheinlichkeit, dass ein Element aus der Urne gezogen wird und somit in die Stichprobe gelangt, ist s/N . Hierbei ist die Ziehung der Kugeln nicht unabhängig voneinander, sondern die Ziehung einer Kugel kann die Ziehungswahrscheinlichkeit einer anderen Kugel vermindern.

Estes schreibt in seinem Artikel von 1950, dass er zuerst an die 1. Variante gedacht hat, aber für die Ableitung der Lernkurve spielt die Art der Ziehung keine Rolle. Ersetze deshalb s/N durch θ (Bower, 1994).

In der SST wird von zwei Reaktionsmöglichkeiten auf zwei Ereignisse ausgegangen. Jedes Merkmal ist mit genau einer Reaktionsmöglichkeit verbunden. Nehmen wir zum Beispiel zwei mögliche Ereignisse, die wir mit E1 und E2 bezeichnen. Die Reaktionen auf die Ereignisse E1 und E2 werden mit A1 und A2 bezeichnet. Ein Beispiel für zwei solche Ereignisse wäre:

- E1: kein Regen
- E2: Regen

Die zugehörigen Reaktionen A1 und A2 auf diese Ereignisse können folgendermaßen aussehen:

- A1: keinen Schirm mitnehmen
- A2: Schirm mitnehmen

Die wahrgenommenen Merkmale, die mit der Reaktion A1 oder A2 verknüpft sein können, können zum Beispiel dunkle Wolken, blauer Himmel, Wind usw. sein. Nehmen wir an, dass alle wahrgenommenen Merkmale mit der Reaktion A2 verknüpft sind, dass aber das Ereignis E1 eintritt und nicht E2, wie erwartet. So werden alle wahrgenommenen Merkmale auf die Reaktion A1 umkonditioniert, da sie vorher mit der Reaktion A2 verbunden waren. Somit ist die Estes Theorie eine Alles oder Nichts Theorie. Wäre statt dem Ereignis E1, dass Ereignis E2 eingetreten, das erwartet wird, so bleiben die wahrgenommenen Merkmale mit der Reaktion A2 verbunden.

Es soll nun die Reaktionswahrscheinlichkeit für die Reaktion A1 bestimmt werden. Mit der Wahrscheinlichkeit p ist ein beliebiges Merkmal mit der Reaktion A1 auf das Ereignis E1 verbunden und mit der Wahrscheinlichkeit $1 - p$ mit der Reaktion A2 auf das Ereignis E2. In jedem Lerndurchgang ändert sich die Wahrscheinlichkeit p . Sei $p(n)$ die Wahrscheinlichkeit im n -ten Lerndurchgang, dann kann $p(n)$ als Wahrscheinlichkeit für eine Reaktion A1 im n -ten Durchgang betrachtet werden. Tritt im n -ten Lerndurchgang das verstärkende Ereignis E1 ein, so können wir $p(n + 1)$ folgendermaßen berechnen (Hilgard & Bower, 1983):

$$p(n + 1) = p(n) + \theta \cdot (1 - p(n)) \quad (4.1)$$

Mit der Wahrscheinlichkeit θ wird ein Merkmal in die Stichprobe aufgenommen. Mit der Wahrscheinlichkeit $p(n)$ sind die Merkmale schon mit der Reaktion A1 auf das Ereignis E1 verknüpft. Bei diesen Merkmalen ändert sich nichts, wenn das Ereignis E1 eintritt. Mit der Wahrscheinlichkeit $1 - p(n)$ sind

die Merkmale nach dem n -ten Durchgang mit der Reaktion A2 auf das Ereignis E2 verknüpft. Tritt nun Ereignis E1 ein, so wird ein Teil der Merkmale, die vorher mit der Reaktion A2 auf das Ereignis E2 verknüpft waren, mit der Reaktion A1 auf das Ereignis E1 verknüpft. Mit der Wahrscheinlichkeit θ sind die mit der Reaktion A2 auf das Ereignis E2 verknüpften Merkmale in die Stichprobe gekommen. Durch Umformung von Gleichung 4.1 erhält man:

$$p(n+1) = (1 - \theta) \cdot p(n) + \theta \quad (4.2)$$

Dies kann man auch folgendermaßen interpretieren (Hilgard & Bower, 1983): Ein Merkmal wird mit der Wahrscheinlichkeit $(1 - \theta)$ nicht in die Stichprobe aufgenommen, d.h sein Zustand $p(n)$ verändert sich nicht. Mit der Wahrscheinlichkeit θ gerät das Merkmal in die Stichprobe und wird mit der Reaktion A1 auf das Ereignis E1 verknüpft. Als nächstes betrachten wir eine Verstärkung des Ereignisses E2 im n -ten Durchgang. Wir können dann $p(n+1)$ folgendermaßen berechnen:

$$p(n+1) = (1 - \theta) \cdot p(n) + \theta \cdot 0 = (1 - \theta) \cdot p(n) \quad (4.3)$$

Mit der Wahrscheinlichkeit $(1 - \theta)$ ist das Merkmal nicht in der Stichprobe enthalten, somit ändert sich die Wahrscheinlichkeit $p(n)$ nicht. Ist das Merkmal in der Stichprobe (mit Wahrscheinlichkeit θ) enthalten, so wird es mit der Reaktion A2 auf das Ereignis E2 verbunden, d.h die Wahrscheinlichkeit, dass es mit der Reaktion A1 auf das Ereignis E1 verbunden wird, ist 0. Zusammengefasst ergeben sich die folgenden beiden Differenzgleichungen, die zusammen ein lineares Lernmodell (Estes Modell) darstellen:

$$p(n+1) = (1 - \theta) \cdot p(n) + \theta, \text{ für E1 im } n\text{-ten Durchgang (Lernen)} \quad (4.4)$$

$$p(n+1) = (1 - \theta) \cdot p(n), \text{ für E2 im } n\text{-ten Durchgang (Interferenz)} \quad (4.5)$$

Wenn immer nur das Ereignis E1 eintritt, so wird in jedem Durchgang Gleichung 4.4 angewendet, d.h. $p(n)$ wird in jedem Schritt größer, solange bis der Grenzwert 1 erreicht wird. Ist $p(n) = 1$, so kann durch Gleichung 4.4 der Wert $p(n)$ nicht mehr verändert werden, d.h. die maximale Assoziationsstärke zur Reaktion A1 ist erreicht. Alle Merkmale sind schon mit der Reaktion A1 auf das Ereignis E1 verbunden. Tritt nun Ereignis E2 ein, so wird die Assoziationsstärke zu der Reaktion A1 durch Gleichung 4.5 vermindert. Der Grenzwert der Gleichung 4.5 ist Null, d.h wenn $p(n) = 0$ ist, dann ist kein Merkmal mehr mit der Reaktion A1 verknüpft.

Tritt im n -ten Durchgang das Null Ereignis E0 ein, so ändert sich die Assoziationsstärke nicht.

$$p(n+1) = p(n) \quad (4.6)$$

Das Null Ereignis bewirkt, dass kein Element aus der Stichprobe mit den Reaktionen A1 oder A2 verbunden wird.

Eine weitere Möglichkeit, die Lerngleichung abzuleiten, wird in (Coombs, Dawes & Tversky, 1970) beschrieben. Hierfür werden folgende Vereinbarungen getroffen:

- s : Anzahl der Elemente in der Stichprobe, sie bleibt während des gesamten Lernvorgangs konstant.
- $C_{s,n}$: Anzahl der konditionierten Elemente in der Stichprobe im n -ten Durchgang.
- N : Anzahl der Stimuluselemente, die wahrgenommen werden können.
- C_n : Anzahl der konditionierten Elemente im n -ten Durchgang.
- p_n : Wahrscheinlichkeit einer A_1 Reaktion im n -ten Durchgang, dies ist das Verhältnis der konditionierten Elemente im n -ten Durchgang zur Anzahl der Stimuluselemente, die wahrgenommen werden können.

Für die Herleitung der Lerngleichung aus dem Estes Modell wird in (Coombs, Dawes & Tversky, 1970) die folgende Annahme gemacht:

Das Verhältnis der Anzahl der konditionierten Elemente insgesamt ist gleich dem Verhältnis der Anzahl der konditionierten Elemente in der Stichprobe.

$$\frac{C_{s,n}}{s} = \frac{C_n}{N} = p_n \quad (4.7)$$

Erfolgt eine Verstärkung, so werden alle Elemente in der Stichprobe mit Wahrscheinlichkeit 1 auf eine A_1 Reaktion konditioniert. Dadurch erhöht sich die Anzahl der mit einer A_1 Reaktion konditionierten Elemente im $(n+1)$ -ten Durchgang um folgende Differenz:

$$C_{n+1} - C_n = s - C_{s,n} = s\left(1 - \frac{C_{s,n}}{s}\right) \quad (4.8)$$

Von den s Elementen in der aktuellen Stichprobe sind $C_{s,n}$ Elemente schon im vorherigen Durchgang auf eine A_1 Reaktion konditioniert. Bei diesen Stimuluselementen ändert sich also nichts. $(s - C_{s,n})$ -Elemente sind noch nicht mit einer A_1 Reaktion verknüpft, somit ist dies der Anteil der Elemente die im $(n+1)$ -ten Durchgang neu mit einer A_1 Reaktion verknüpft werden.

Wird die obere Gleichung durch N dividiert und werden danach die Terme umgestellt, so erhält man folgende Gleichung:

$$\frac{C_{n+1}}{N} = \frac{C_n}{N} + \frac{s}{N} \left(1 - \frac{C_{s,n}}{s}\right) \quad (4.9)$$

Setze $\theta = \frac{s}{N}$ und wegen 4.7 folgt:

$$p_{n+1} = p_n + \theta(1 - p_n) \quad (4.10)$$

Dies ist die Lerngleichung aus dem Estes Modell, die schon vorher beschrieben wurde, nur etwas anders hergeleitet. Die Gleichung für Interferenz lässt sich entsprechend herleiten.

Das Estes Modell wurde unter anderem an Experimenten zum Lernen von Wahrscheinlichkeiten überprüft. Bei diesen Versuchen haben die Teilnehmer die Aufgabe vorausszusagen, welches von zwei möglichen Ereignissen im nächsten Durchgang eintritt. Nach der Voraussage tritt eines der beiden Ereignisse ein. Welches Ereignis eintritt, wird bei diesen Versuchen zufällig ausgewählt, so dass die Teilnehmer keine Information haben mit der sie das nächste Ereignis sicher voraussagen können.

Tritt das verstärkende Ereignis E1 mit einer konstanten Wahrscheinlichkeit von π ein und das verstärkende Ereignis E2 mit einer Wahrscheinlichkeit von $(1 - \pi)$, so erhalten wir die Wahrscheinlichkeit für ein A1 Reaktion im $(n+1)$ -ten Durchgang $p(n+1)$ durch multiplizieren der Gleichung 4.4 mit π und der Gleichung 4.5 mit $(1 - \pi)$.

$$p(n+1) = \pi[(1-\theta) \cdot p(n) + \theta] + (1-\pi)[(1-\theta) \cdot p(n)] = (1-\theta) \cdot p(n) + \theta \cdot \pi \quad (4.11)$$

Die obige lineare Differenzgleichung hat folgende Lösung (Foppa, 1965).

$$p(n) = \pi - (\pi - p(0)) \cdot (1 - \theta)^n \quad (0 < \theta < 1) \quad (4.12)$$

Da θ nur Werte annehmen kann, die kleiner als eins sind, wird der Term $(1 - \theta)^n$ für wachsendes n immer kleiner. Somit hat die Gleichung 4.12 folgenden Grenzwert:

$$\lim_{n \rightarrow \infty} [\pi - (\pi - p(0)) \cdot (1 - \theta)^n] = \pi \quad (4.13)$$

Der Grenzwert der Gleichung 4.12 ist die Wahrscheinlichkeit, dass eine Reaktion A1 verstärkt wird. Gleichung 4.12 nennt man auch *probability-matching prediction*. Sie wurde in vielen Experimenten mit Menschen und Tieren getestet und bestätigt.

Verschiedene Theorien zum Kurzzeitgedächtnis und zum freien Erinnern (free recall) sind von Estes beeinflusst worden (Bower, 1994). Es existiert eine Festschrift zu Ehren von Estes, die insgesamt 21 Kapitel von Autoren enthält, dessen Arbeiten von den Theorien von Estes beeinflusst wurden (Healy, Kosslyn & Shiffrin, 1992). Dies sind unter anderem das SAM Modell von Raaijmakers & Shiffrin (Raaijmakers & Shiffrin, 1980) und das Kurzzeitgedächtnis Modell von Murdock (Murdock, 1974).

Das PASS Modell von Sedlmeier (Sedlmeier, 1999), das zur Voraussage von Häufigkeitsschätzungen entwickelt wurde, ist eine Weiterentwicklung des Estes Modells.

Das Estes Modell wurde nicht speziell als Modell für die klassische Konditionierung entwickelt, wie das Rescorla & Wagner Modell, das im nächsten Abschnitt beschrieben werden soll.

4.3 Das Rescorla & Wagner Modell

Nach der Theorie des bedingten Reflex soll die Stärke eines bedingten Reflexes von der relativen Häufigkeit abhängen, dass vor dem US der CS auftritt. Andererseits zeigt sich in den Konditionierungsexperimenten mit mehrere Reizen, dass die Stärke eines bedingten Reflexes auch von den anderen im Experiment dargebotenen Reizen abhängt. Beispiele hierfür sind die Experimente zu Blocking und zur Überschattung (Overshadowing) von Reizen.

In einem Blocking Experiment erfolgt zuerst eine Konditionierung mit einem CS_1 . Nachdem eine Verbindung zwischen CS_1 und US aufgebaut worden ist, wird ein weiterer neutraler Reiz zusätzlich zu dem CS_1 dargeboten, d.h. immer wenn der CS_1 und der US auftritt, wird auch der neutrale Reiz ebenfalls dargeboten. Zuerst erfolgt zum Beispiel eine Konditionierung von Glocke auf Futter. Wenn der Hund nach dem Glockenton mit vermehrter Speichelsekretion reagiert, erfolgt eine Konditionierung von Glocke und Licht auf den US Futter. Wird jetzt nur das Licht ohne die Glocke dargeboten, so zeigt der Hund keine Reaktion, wird aber die Glocke als Reiz verwendet, so führt dies bei dem Hund zu vermehrter Speichelsekretion. Eine Verbindung zwischen Licht und Futter wird nicht erlernt.

Dieses Phänomen kann mit dem Estes Modell nicht erklärt werden, da bei der Berechnung der Assoziationsstärke immer nur ein Reiz berücksichtigt wird. Von Rescorla & Wagner wurde deshalb ein mathematisches Lernmodell, indem die Assoziationsstärke zwischen CS und US auch von den anderen dargebotenen Reizen abhängt, entwickelt. Die Veränderung der Assoziationsstärke ΔV_{ij} zwischen einem CS_i und einem US_j wird in dem Rescorla & Wagner Modell nach folgender Gleichung berechnet:

$$\Delta V_{ij} = \alpha_i \beta_j (\lambda_j - \sum_k V_{kj}) \quad (4.14)$$

Mit α_i und β_j werden die beiden Lernparameter bezeichnet, die von der Reizstärke (Salienz) des jeweiligen CS und des jeweiligen US abhängen. Die Werte von α_i und β_j können zwischen 0 und 1 variieren. Je näher die Werte von α_i und β_j sich dem Wert 1 annähern, je schneller wird die maximale Assoziationsstärke erreicht. Diese wird mit λ_j bezeichnet. Üblicherweise wird $\lambda_j = 1$ gewählt, falls der US_j im aktuellen Lerndurchgang auftritt.

Entsprechend wird $\lambda_j = 0$ gewählt, falls der US_j im aktuellen Lerndurchgang nicht auftritt. In der Summe $\sum_k V_{kj}$ werden alle Assoziationsstärken zwischen dem US j und allen anderen Reizen CS_k , die im aktuellen Lerndurchgang präsent sind, aufaddiert. Wird in einem Konditionierungsexperiment nur ein CS verwendet, so erhält man folgende vereinfachte Regel für das Rescorla & Wagner Modell:

$$\Delta V_{ij} = \alpha_i \beta_j (\lambda_j - V_{ij}) \quad (4.15)$$

Werden die beiden Parameter für Lernen α_i und β_j zu einem Parameter ($\theta = \alpha_i \beta_j$) zusammengefasst, so erhält man das Estes Modell in einer Gleichung zusammengefasst.

$$\Delta V_{ij} = \theta (\lambda_j - V_{ij}) \quad (4.16)$$

Für $\lambda_j = 1$ erhält man aus der obigen Gleichung die Lernregel vom Estes Modell:

$$\Delta V_{ij} = \theta (1 - V_{ij}) \quad (4.17)$$

Für $\lambda_j = 0$ erhält man entsprechend die Interferenzregel vom Estes Modell:

$$\Delta V_{ij} = \theta (0 - V_{ij}) = \theta \cdot V_{ij} \quad (4.18)$$

Somit ist das Estes Modell ein Spezialfall des Rescorla & Wagner Modells. Bei dem Estes Modell hängt die Assoziationsstärke von der Wahrscheinlichkeit ab, dass der US unmittelbar nach dem CS auftritt (Kontiguität). Beim Rescorla & Wagner Modell ist die Kontingenz entscheidend für die Konditionierung. Die Kontingenz zwischen CS und US ist die Differenz zwischen der Wahrscheinlichkeit, dass der US auftritt, wenn zuvor der CS aufgetreten ist und der Wahrscheinlichkeit, dass der US auftritt, wenn der CS nicht aufgetreten ist.

In einem Experiment, in dem ein CS und ein US vorkommt, gibt es vier Kombinationsmöglichkeiten für diese:

1. Der CS wird dargeboten und darauf folgt der US.
2. Der CS wird gegeben, es folgt aber kein US.
3. Der US tritt ohne den CS auf.
4. Weder CS noch US treten auf.

In einer Vier-Felder Tafel oder Kontingenzmatrix werden die Wahrscheinlichkeiten für die vier Kombinationsmöglichkeiten angegeben (siehe Tabelle 4.1).

	US	\overline{US}	Summe
CS	a	b	a + b
\overline{CS}	c	d	c + d

Tabelle 4.1: Kontingenzmatrix für einen CS und einen US

Die Kontingenz zwischen CS und US (ΔP) wird folgendermaßen berechnet:

$$\begin{aligned} \Delta P &= P(US | CS) - P(US | \overline{CS}) = \frac{P(US \cap CS)}{P(CS)} - \frac{P(US \cap \overline{CS})}{P(\overline{CS})} \\ &= \frac{a}{a+b} - \frac{c}{c+d} \end{aligned} \quad (4.19)$$

Chapman und Robbins (Chapman & Robbins, 1990) beschreiben, wie mit Hilfe des Rescorla & Wagner Modells die Kontingenz zwischen CS und US ermittelt werden kann. Dazu machen sie die Annahme, dass ein CS nie allein auftritt, sondern immer in einem Kontext von einem anderen CS, dem sogenannten Hintergrund, eingebettet ist. Dieser Hintergrund kann zum Beispiel der Versuchsraum sein, in dem das Experiment durchgeführt wird. Der Hintergrund ist während des gesamten Experimentes präsent. Das bedeutet, wenn der CS gegeben wird, sind zwei Reize aktiv, und zwar der Hintergrund als CS und der eigentliche CS. Der Hintergrund als CS wird im folgenden mit X bezeichnet und der eigentliche CS mit A. Der CS ist somit ein kombinierter Reiz, der aus den Reizen A und X besteht. Wird der CS A nicht präsentiert, so ist immer noch der CS X präsent. Für die obige Vier-Felder

	US	\overline{US}	Summe
AX	a	b	a + b
X	c	d	c + d

Tabelle 4.2: Vier-Felder Tafel für kombinierten Reiz AX und Reiz X allein

Tafel 4.1 bedeutet dies, dass der CS durch XA und \overline{CS} durch A ersetzt wird (siehe Tabelle 4.2).

Wird der CS A präsentiert und folgt darauf der US, so ergeben sich beim Rescorla & Wagner Modell folgende Gleichungen:

$$\Delta V_A = \alpha\beta(1 - (V_X + V_A)) \quad (4.20)$$

$$\Delta V_X = \alpha\beta(1 - (V_X + V_A)) \quad (4.21)$$

In der ersten Gleichung wird die Veränderung der Assoziationsstärke für den CS A berechnet, in der zweiten Gleichung wird die Veränderung der Assoziationsstärke für den Hintergrund X berechnet. λ ist in beiden Gleichungen gleich 1, da der US aufgetreten ist.

Wird der CS A präsentiert und folgt darauf kein US, so ist λ gleich 0, und die Assoziationsstärken von dem CS A und dem Hintergrund werden vermindert.

$$\Delta V_A = \alpha\beta(0 - (V_X + V_A)) = -\alpha\beta(V_X + V_A) \quad (4.22)$$

$$\Delta V_X = \alpha\beta(0 - (V_X + V_A)) = -\alpha\beta(V_X + V_A) \quad (4.23)$$

Tritt der US ohne den CS auf, so wird die Assoziationsstärke für den Hintergrund erhöht, weil der Hintergrund als CS während des gesamten Experimentes aktiv ist. Da der CS A nicht aktiv ist, wird seine Assoziationsstärke nicht verändert. Bei der Berechnung der Assoziationsstärke für den Hintergrund wird die Assoziationsstärke des CS A nicht berücksichtigt.

$$\Delta V_X = \alpha\beta(1 - V_X) \quad (4.24)$$

Wird der CS A nicht gegeben und tritt kein US auf, so wird die Assoziationsstärke für den Hintergrund vermindert, da dieser CS aktiv ist, aber kein US folgt.

$$\Delta V_X = \alpha\beta(0 - V_X) = -\alpha\beta V_X \quad (4.25)$$

Die mittlere Änderung von V_A ergibt sich, wenn die obigen Gleichungen, in denen V_A berechnet wird, erst mit der relativen Häufigkeit ihrer Anwendung gewichtet und anschließend summiert werden. Wenn der CS A präsentiert wird und der US auftritt, wird die Assoziationsstärke von V_A gemäß Gleichung 4.20 verändert. Dieser Fall tritt laut der Vier - Felder Tafel, die in der Tabelle 4.2 dargestellt wird, a mal auf, d.h die Gleichung 4.20 wird mit a multipliziert. Die Assoziationsstärke V_A wird im Fall, CS A wird präsentiert, aber kein US, ebenfalls verändert, und zwar gemäß Gleichung 4.22. Laut Vier - Felder Tafel tritt der Fall, CS A wird präsentiert, aber kein US, b mal ein, d.h Gleichung 4.22 wird mit b multipliziert. Somit ergibt sich folgende mittlere Veränderung für ΔV_A :

$$\text{Mean}\Delta V_A = a\alpha\beta(1 - (V_X + V_A)) + b\alpha\beta(0 - (V_X + V_A)) \quad (4.26)$$

Die Assoziationsstärke V_X muss in jedem der vier Fälle berechnet werden, weil der Hintergrund als CS während des gesamten Experimentes aktiv ist.

$$\begin{aligned} \text{Mean}\Delta V_X = a\alpha\beta(1 - (V_X + V_A)) + b\alpha\beta(0 - (V_X + V_A)) + c\alpha\beta(1 - V_X) \\ + d\alpha\beta(0 - V_X) \end{aligned} \quad (4.27)$$

Beide Gleichungen können vereinfacht werden, indem durch die Parameter α und β dividiert wird und anschließend die Terme ausmultipliziert werden. Die beiden Gleichungen haben dann folgende Form:

$$\frac{\text{Mean}\Delta V_A}{\alpha\beta} = a - V_A(a + b) - V_X(a + b) \quad (4.28)$$

$$\frac{\text{Mean}\Delta V_X}{\alpha\beta} = a - V_A(a + b) + c - V_X(a + b + c + d) \quad (4.29)$$

Wenn ΔV_A und ΔV_X beide Null werden, ist die maximale Assoziationsstärke erreicht und es findet kein Lernen mehr statt. Ab dieser Stelle können die Assoziationsstärken von A und X als konstant betrachtet werden. Gesucht ist die Assoziationsstärke des Reizes A. Um diese zu bestimmen, werden die beiden obigen Gleichungen gleich Null gesetzt.

$$a - V_A(a + b) - V_X(a + b) = 0 \quad (4.30)$$

$$a - V_A(a + b) + c - V_X(a + b + c + d) = 0 \quad (4.31)$$

Die Gleichung 4.31 wird nach V_X aufgelöst.

$$V_X = \frac{a + c - V_A(a + b)}{a + b + c + d} \quad (4.32)$$

Anschließend wird V_X in die Gleichung 4.30 eingesetzt, um die gesuchte Assoziationsstärke des Reizes A zu bekommen.

$$a - V_A(a + b) - \frac{a + c - V_A(a + b)}{a + b + c + d}(a + b) = 0 \quad (4.33)$$

$$V_A = \frac{a}{a + b} - \frac{c}{c + d} \quad (4.34)$$

In den folgenden Abschnitten wird beschrieben, wie drei Phänomene, die in klassischen Konditionierungsexperimenten mit mehreren Reizen beobachtet wurden, mit Hilfe des Rescorla & Wagner Modells erklärt werden können.

4.3.1 Blocking

Das Rescorla & Wagner Modell kann den Blockingeffekt folgendermaßen erklären: In der ersten Phase wird eine Assoziation zwischen dem CS_1 und dem US aufgebaut bis die maximale Assoziationsstärke λ erreicht worden ist, d.h. $V_{CS_1,US} = \lambda$. Für den zweiten CS_2 gilt: $V_{CS_2,US} = 0$, da er in dieser Phase des Experimentes nicht auftritt. In der zweiten Phase des Experimentes wird vor dem Auftreten des US der zweite CS_2 zusätzlich zu dem ersten CS präsentiert. Für den CS_1 ist die maximale Assoziationsstärke schon erreicht, sie kann nicht weiter ansteigen. Für den zweiten CS_2 bleibt die Assoziationsstärke gleich Null, da beim Lernen von der maximal möglichen Assoziationsstärke (λ) die Summe aller Assoziationsstärken vom US zum CS, $V_{CS_1,US} + V_{CS_2,US}$, subtrahiert wird. Da aber die Assoziationsstärke $V_{CS_1,US} = \lambda$ ist, ist schon der Maximalwert erreicht. Die Änderung der Assoziationsstärke von CS_2 und dem US $\Delta V_{CS_2,US}$ hat den Wert Null. Nach dem Estes Modell dagegen wird eine Assoziation zwischen CS_2 und dem US gelernt, da die Assoziationsstärke von CS_2 und dem US unabhängig von der Assoziationsstärke von CS_1 und dem US ist.

4.3.2 Overshadowing (Überschattung von Reizen)

Ein weiteres Phänomen, das beim klassischen Konditionieren auftritt, wenn kombinierte Reize auf einen US konditioniert werden, ist die Überschattung (Overshadowing) von Reizen. In einem Überschattungsexperiment werden zwei neutrale Reize CS_1 und CS_2 auf einen US konditioniert. Zum Beispiel

nach Ton CS_1 und Licht CS_2 als kombinierter Reiz wird der US Futter gegeben. Wird dies einige Male wiederholt, so ergibt ein separater Test von CS_1 oder CS_2 , dass die konditionierte Reaktion Speichelfluss bei einem der beiden Reize größer ist als bei dem anderen Reiz. Der Reiz, der zu mehr Speichelfluss führt, überschattet den anderen Reiz, der nur zu einem geringen Speichelfluss führt.

Im Rescorla & Wagner Modell wird ein Überschattungsexperiment dadurch simuliert, dass dem überschattenden Reiz eine größere Saliens zuge-wiesen wird, als dem Reiz, der überschattet wird. Z.B. $\alpha_1 = 0.7$ für den überschattenden Reiz CS_1 und $\alpha_2 = 0.3$ für den Reiz CS_2 , der überschattet wird. Während der Konditionierung von CS_1 und CS_2 auf den US steigt die Assoziationsstärke zwischen dem CS_1 und dem US stärker an als die Assoziationsstärke zwischen dem CS_2 und dem US. Sobald beide Assoziationsstärken zusammen die maximal mögliche Assoziationsstärke erreicht haben, findet kein Lernen mehr statt ($V_{CS_1,US} + V_{CS_2,US} = \lambda$). Die Assoziationsstärke $V_{CS_1,US}$ ist größer als $V_{CS_2,US}$.

4.3.3 Inhibition

Ein weiteres Phänomen, das vom Rescorla & Wagner Modell erklärt werden kann, ist die konditionierte Hemmung (Inhibition). Bei der konditionierten Hemmung wird in der ersten Phase ein CS_1 auf einen US konditioniert. In der zweiten Phase wird ein zweiter Reiz CS_2 zusammen mit dem Reiz CS_1 dargeboten. Dieser zweite Reiz CS_2 ist vorher noch nicht zusammen mit dem US aufgetreten. Seine Assoziationsstärke zum US ist somit 0. Somit gilt für die beiden Assoziationsstärken: $V_{CS_1,US} + V_{CS_2,US} = V_{CS_1,US}$. Nach der Darbietung von CS_1 und CS_2 folgt aber kein US, somit ist λ gleich 0, d. h. die Verbindungen der Reize CS_1 und CS_2 zum US werden gelöscht. Da aber der Reiz CS_2 noch gar keine Assoziationsstärke zum US aufbauen konnte, wird seine Assoziationsstärke zum US negativ. Der Reiz CS_2 wird zu einem inhibitorischen Reiz.

4.4 Das Gallistel Modell - Rate Estimation Theorie (RET)

Bei den Lerntheorien von Estes und Rescorla & Wagner wird das Auftreten von vier möglichen Ereignissen als diskretes Ereignis betrachtet. Gallistel untersucht dagegen wie lange ein Reiz aktiv ist und wie häufig der US auftritt, während der Reiz aktiv ist, oder wie viele Zeiteinheiten verstreichen bis der US auftritt. Bei den Lerntheorien von Estes und Rescorla & Wagner wird

nicht berücksichtigt, wie viel Zeit zwischen dem Auftreten des CS und dem Auftreten des US vergeht. Die Darbietungsdauer des CS wird ebenfalls nicht berücksichtigt. Deswegen muss bei den Lerntheorien von Estes und Rescorla & Wagner das Experiment getaktet werden, um festzulegen, wann ein Lern-durchgang beendet ist. Dies ist bei der RET Theorie nicht nötig, da nur die Auftretensrate des US während der Darbietungsdauer des CS und die Darbietungsdauer des CS als Information benötigt wird.

Gallistel bildet das Verhältnis von Auftretenshäufigkeit des US zur Darbietungsdauer des CS. Aus diesem Verhältnis soll bestimmt werden, wie gut ein CS einen US voraussagt. Die Bestimmung dieses Verhältnisses allein reicht aber nicht aus. Denn laut Gallistel existieren in einem Konditionierungsexperiment mindestens zwei Reize, da der Hintergrund auch als Reiz angesehen wird, der während des gesamten Experimentes aktiv ist (Gallistel, 1990). Sind mehrere Reize zur gleichen Zeit aktiv und tritt der US auf, so kann der US nicht eindeutig einem Reiz zugeordnet werden. Welchem Reiz der US zugeordnet wird, ist nach Gallistel das Ergebnis eines sogenannten Partitionierungsprozesses (Gallistel, 1990). Dabei wird für jeden Reiz eine Gleichung aufgestellt. Es wird das Verhältnis von der Auftretenshäufigkeit des US zur Darbietungsdauer des CS gebildet. Nach Gallistel (Gallistel, 1990) ist dieses Verhältnis das unkorrigierte Verhältnis für die Schätzung der Auftretensrate des US zur Darbietungsdauer des CS. Dieses Verhältnis wird als unkorrigiertes Verhältnis bezeichnet, da nur ein CS betrachtet wird und nicht die anderen Reize, die auch zur selben Zeit des aktuell betrachteten CS aktiv sind. Um das korrekte Verhältnis zu ermitteln, muss ein lineares Gleichungssystem, das für jeden CS aus einer Gleichung besteht, gelöst werden. Für zwei Reize besteht dieses Gleichungssystem aus zwei Gleichungen mit zwei Unbekannten. Die Unbekannten in diesem Gleichungssystem sind die korrekten Verhältnisse, die bestimmt werden sollen. Für zwei Reize und einen US ergibt sich folgendes Gleichungssystem (Gallistel, 1990):

$$\frac{N_1}{T_1} = \lambda_1 + \frac{T_{12}}{T_1} \lambda_2 \quad (4.35)$$

$$\frac{N_2}{T_2} = \frac{T_{21}}{T_2} \lambda_1 + \lambda_2 \quad (4.36)$$

Mit N_1 wird die Auftretenshäufigkeit des US, während der CS_1 aktiv ist, bezeichnet. Entsprechend ist N_2 die Auftretenshäufigkeit des US, wenn der Reiz CS_2 dargeboten wird. T_1 ist die Darbietungsdauer des CS_1 . Dies ist die kumulierte Zeit, die der CS_1 präsentiert wird. Mit $T_{1,2}$ wird die Zeitspanne bezeichnet, in der sowohl CS_1 als auch CS_2 dargeboten werden. Dies ist also die Zeitspanne, in der sich die beiden Reize überlappen. Die Unbekannte

λ_1 steht für die korrekte Schätzung der Auftretensrate des US, wenn der CS_1 dargeboten wird. Entsprechend gibt λ_2 das korrekte Verhältnis für die Auftretensrate des US an, wenn der CS_2 dargeboten wird.

In einem Verzögerungskonditionierungsexperiment (delay conditioning) wird mit T_i die Wartezeit auf die Verstärkung (reinforcement latency) von einem CS_i bezeichnet.

Die obigen Gleichungen besagen, dass das unkorrigierte Verhältnis von US Auftretenshäufigkeit zur Darbietungsdauer des betrachteten CS gleich dem korrigierten Verhältnis plus einem Fehlerterm ist. Dieser Fehlerterm entsteht durch die Zeitintervalle, in denen beide Reize präsentiert werden. Wird gerade der CS_1 betrachtet, so entsteht der Fehler dadurch, dass während der CS_1 dargeboten wird noch zusätzlich der CS_2 dargeboten wird. Bei dem unkorrigierten Verhältnis wird jeweils nur ein CS berücksichtigt, d.h. es wird so getan, als ob es den anderen Reiz gar nicht gäbe. Es entsteht ein Fehlerterm, wenn beide Reize gemeinsam auftreten. Der Fehlerterm für CS_1 besteht aus dem korrekten Verhältnis für den CS_2 multipliziert mit dem gemeinsamen Zeitintervall von CS_1 und CS_2 dividiert durch die Darbietungsdauer des CS_1 . Die korrigierten Verhältnisse erfüllen das additive Prinzip, d.h. die Auftretenshäufigkeiten des US, während mehrere Reize aktiv sind, ist gleich der Summe der korrigierten Verhältnisse, die den einzelnen Reizen zugeordnet werden. Zur Bestimmung der korrekten Raten für die einzelnen Reize müssen die obigen beiden Gleichungen jeweils nach den Unbekannten λ_1 und λ_2 aufgelöst werden.

Für n Reize erhält man ein Gleichungssystem mit n Gleichungen und n Variablen $\lambda_1, \lambda_2, \dots, \lambda_n$. Mit Hilfe von Matrizenalgebra lässt sich ein solches Gleichungssystem lösen. Sei λ_u der Vektor mit den unkorrigierten Raten, dieser hat dann folgende Form:

$$\lambda_u = \begin{pmatrix} \frac{N_1}{T_1} \\ \frac{N_2}{T_2} \\ \vdots \\ \vdots \\ \frac{N_n}{T_n} \end{pmatrix} \quad (4.37)$$

In der Matrix T wird die gemeinsame Darbietungsdauer von zwei Reizen im Verhältnis zur Darbietungsdauer des betrachteten CS gespeichert.

$$T = \begin{pmatrix} 1 & \frac{T_{12}}{T_1} & \dots & \frac{T_{1n}}{T_1} \\ \frac{T_{21}}{T_2} & 1 & \dots & \frac{T_{2n}}{T_2} \\ \vdots & & & \\ \vdots & & & \\ \frac{T_{n1}}{T_n} & \frac{T_{n2}}{T_n} & \dots & 1 \end{pmatrix} \quad (4.38)$$

In der ersten Zeile der Matrix T stehen die gemeinsamen Präsenzzeiten des CS_1 mit den anderen Reizen dividiert durch die gesamte Präsenzzeit von CS_1 . Die 1 in der Diagonale ergibt sich aus ($T_{ii} = T_i$). Der Vektor mit den korrigierten Raten wird mit λ_c bezeichnet.

$$\lambda_c = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \vdots \\ \lambda_n \end{pmatrix} \quad (4.39)$$

Das zu lösende Gleichungssystem hat in der Matrizen Schreibweise folgende Gestalt:

$$\lambda_u = T \cdot \lambda_c \quad (4.40)$$

Gesucht ist der Vektor λ_c mit den korrekten Raten. Diesen erhält man, indem man die obige Gleichung mit der inversen Matrix von T multipliziert.

$$\lambda_c = T^{-1} \cdot \lambda_u \quad (4.41)$$

Wenn die Determinante der Matrix T gleich Null ist, so hat das Gleichungssystem unendlich viele Lösungen. Dies bedeutet, dass das Gleichungssystem redundante Reize enthält. Es existieren nicht genügend unabhängige Gleichungen für die zu schätzenden Raten. Dieses Problem lässt sich lösen, indem man ein oder mehrere Reize ignoriert, so dass die Determinante ungleich Null wird. Durch das Entfernen eines Reizes entstehen natürlich auch unterschiedliche Lösungen. Nach Gallistel (Gallistel & Gibbon, 2000) ist die Lösung, bei der die Summe der Absolutwerte des Vektors λ_c minimal wird, die optimale Lösung.

$$\sum_{i=1}^n |\lambda_i| = \min \quad (4.42)$$

In ihrer Veröffentlichung (Gallistel & Gibbon, 2000) präsentieren Gallistel und Gibbon eine kognitive Theorie, die die RET Theorie enthält. Ihre Theorie

besteht aus SET (Scalar Expectancy Theory) und RET. SET wurde von Gibbon entwickelt. Die SET Theorie sagt voraus, wann eine konditionierte Reaktion erfolgt und wie lange die konditionierte Reaktion anhält. Mit Hilfe der RET Theorie wird entschieden, ob überhaupt eine Reaktion erfolgen soll.

In den folgenden Abschnitten werden einige Phänomene, die die RET Theorie erklären kann, beschrieben. Sie sind aus dem Buch von Gallistel (Gallistel, 1990) entnommen.

4.4.1 Background Conditioning

Als *Background Conditioning* Experiment bezeichnet man den Prozess, dass die während des Experimentes bestehende Situation als CS mit den experimentell kontrollierten Stimuli verknüpft wird. In einem *Background Conditioning* Experiment ist der Hintergrund (z.B. der Raum, in dem das Experiment stattfindet) als CS während des gesamten Experimentes aktiv. Bezeichnet man den Hintergrund mit CS_1 und einen anderen beliebigen Reiz mit CS_2 , so gilt für das gemeinsame Zeitintervall: $T_{12} = T_2$. Für die beiden Reize ergeben sich dann folgende zwei Gleichungen (Gallistel, 1990):

$$\frac{N_1}{T_1} = \lambda_1 + \frac{T_2}{T_1} \lambda_2 \quad (4.43)$$

$$\frac{N_2}{T_2} = \lambda_1 + \lambda_2 \quad (4.44)$$

Auflösen nach λ_2 ergibt den folgenden Term:

$$\lambda_2 = \frac{\frac{N_2}{T_2} - \frac{N_1}{T_1}}{1 - \frac{T_2}{T_1}} \quad (4.45)$$

Der Zähler wird Null, wenn $\frac{N_2}{T_2} = \frac{N_1}{T_1}$, d.h. das Verhältnis der Auftretenshäufigkeit des US zur Darbietungsdauer des CS_1 ist gleich dem Verhältnis der Auftretenshäufigkeit des US zur Darbietungsdauer des CS_2 . In dem folgenden typischen *Background Conditioning* Experiment tritt genau dieser Fall ein. Ein Versuchstier sitzt 50 min in einem Experimentierkäfig, d.h. $T_1 = 50$. Ein US tritt insgesamt 10 mal auf, während das Tier in dem Käfig sitzt, d.h. $N_1 = 10$. Von diesen 10 mal tritt der US zwei mal auf, wenn ein Ton CS_2 zu hören ist, d.h. $N_2 = 2$. Der Ton ist insgesamt 10 Minuten zu hören, d.h. $T_2 = 10$. Das Tier reagiert im Experiment nicht auf den CS_2 und genau dies wird durch die obige Gleichung auch vorausgesagt. Der US wird nicht dem CS_2 zugeordnet (Gallistel, 1990).

Ändert man das obige Experiment so, dass der US nur auftritt, wenn der CS_2 aktiv ist, d.h. der Ton zu hören ist, so ist $N_1 = N_2$.

Wird $N_1 = N_2$ in die obige Gleichung eingesetzt, so erhält man:

$$\lambda_2 = \frac{\frac{N_1}{T_2} - \frac{N_1}{T_1}}{1 - \frac{T_2}{T_1}} = \frac{\frac{N_1 \cdot T_1 - N_1 \cdot T_2}{T_2 \cdot T_1}}{\frac{T_1 - T_2}{T_1}} = \frac{T_1}{T_1 - T_2} \cdot \frac{N_1 \cdot (T_1 - T_2)}{T_1 \cdot T_2} = \frac{N_1}{T_2} \quad (4.46)$$

Die obige Gleichung besagt, dass die US Auftretensrate, die dem CS_2 zugeordnet wird, die Anzahl der beobachteten unkonditionierten Stimuli ist, dividiert durch die Gesamtdauer. Das Auftreten des US in diesem Experiment wird dem Ton zugeordnet.

4.4.2 Blocking

In einem typischen *Blocking* Experiment sind drei Reize präsent, einmal der Hintergrund (Experimentierkäfig oder Versuchsraum) als CS_1 und zwei weitere Reize CS_2 und CS_3 . Der CS_2 wird immer dann dargeboten, wenn auch der CS_3 dargeboten wird. Gallistel bezeichnet den Reiz CS_2 deswegen auch als lokalen Hintergrund (Gallistel, 1990). Somit kann das *Blocking* Experiment auch als ein *Background Conditioning* Experiment betrachtet werden. Der CS_3 ist der blockierte Reiz oder der lokale Hintergrund. Für ein Blocking Experiment gelten folgende Bedingungen (Gallistel, 1990):

Der US tritt nur auf, wenn der CS_2 aktiv ist: $N_2 = N_1$.

CS_1 und CS_2 sind gemeinsam aktiv, genau dann wenn der CS_2 aktiv ist, d.h. $T_{12} = T_2$.

CS_2 und CS_3 sind gemeinsam aktiv, genau dann wenn der CS_3 aktiv ist, d.h. $T_{23} = T_3$.

CS_1 und CS_3 sind gemeinsam aktiv, genau dann wenn der CS_3 aktiv ist, d.h. $T_{13} = T_3$.

Das Verhältnis der US Auftretensrate, wenn der CS_3 aktiv ist, dividiert durch die Auftretensrate des US, wenn der CS_2 aktiv ist, ist gleich dem Verhältnis von der Darbietungsdauer des CS_3 zur Darbietungsdauer des CS_2 .

$$\frac{N_3}{N_2} = \frac{T_3}{T_2} \quad (4.47)$$

Diese Bedingungen werden in das folgende Gleichungssystem für das *Blocking* Experiment eingesetzt.

$$\frac{N_1}{T_1} = \lambda_1 + \frac{T_{12}}{T_1} \lambda_2 + \frac{T_{13}}{T_1} \lambda_3 \quad (4.48)$$

$$\frac{N_2}{T_2} = \frac{T_{12}}{T_2} \lambda_1 + \lambda_2 + \frac{T_{23}}{T_2} \lambda_3 \quad (4.49)$$

$$\frac{N_3}{T_3} = \frac{T_{13}}{T_3}\lambda_1 + \frac{T_{23}}{T_3}\lambda_2 + \lambda_3 \quad (4.50)$$

Nach dem Einsetzen der obigen Bedingungen und dem Auflösen nach λ_1 , λ_2 und λ_3 erhalten wir:

$$\lambda_1 = \frac{N_2}{T_1} - \frac{T_2}{T_1}\lambda_2 - \frac{T_3}{T_1}\lambda_3 \quad (4.51)$$

$$\lambda_2 = \frac{N_2}{T_2} - \lambda_1 - \frac{T_3}{T_2}\lambda_3 \quad (4.52)$$

$$\lambda_3 = \frac{N_2}{T_2} - \lambda_1 - \lambda_2 \quad (4.53)$$

Ein Gleichungssystem ist genau dann eindeutig lösbar, wenn seine Determinante ungleich Null ist, sonst hat das Gleichungssystem unendlich viele Lösungen. Die Matrix T hat für das obige Gleichungssystem folgende Form:

$$T = \begin{pmatrix} 1 & \frac{T_2}{T_1} & \frac{T_3}{T_1} \\ 1 & 1 & \frac{T_3}{T_2} \\ 1 & 1 & 1 \end{pmatrix} \quad (4.54)$$

Um die Determinante von T zu bestimmen, wird nach der dritten Zeile entwickelt:

$$\det(T) = \det \begin{pmatrix} \frac{T_2}{T_1} & \frac{T_3}{T_1} \\ 1 & \frac{T_3}{T_2} \end{pmatrix} - \det \begin{pmatrix} 1 & \frac{T_3}{T_1} \\ 1 & \frac{T_3}{T_2} \end{pmatrix} + \det \begin{pmatrix} 1 & \frac{T_2}{T_1} \\ 1 & 1 \end{pmatrix} \quad (4.55)$$

$$\det(T) = \frac{T_2}{T_1} \cdot \frac{T_3}{T_2} - \frac{T_3}{T_1} - 1 \left(\frac{T_3}{T_2} - \frac{T_3}{T_1} \right) + 1 - \frac{T_2}{T_1} = 1 - \frac{T_2}{T_1} - \frac{T_3}{T_2} + \frac{T_3}{T_1} \quad (4.56)$$

Die Determinante von T ist ungleich Null, wenn $(T_3 < T_2 < T_1)$ erfüllt ist. Diese Bedingung ist in dem *Blocking* Experiment erfüllt.

Eine Lösung für das Gleichungssystem ist $\lambda_1 = 0$, $\lambda_2 = \frac{N_2}{T_2}$, $\lambda_3 = 0$ (Gallistel, 1990). Dies kann man leicht durch das Einsetzen in das obige Gleichungssystem nachprüfen. In dem *Blocking* Experiment wird das Auftreten des US nur dem CS_2 zugeordnet, nicht aber dem CS_3 oder dem CS_1 . Da kein US während der Darbietung von CS_1 auftritt, kann man das Gleichungssystem vereinfachen, indem man die Gleichung für den CS_1 streicht. Dann unterscheidet sich die mathematische Analyse vom *Blocking* Experiment nicht vom *Background* Experiment (Gallistel, 1990). *Blocking* beim Gallistel Modell unterscheidet sich vom *Blocking* beim Rescorla & Wagner Modell. Beim

Rescorla & Wagner Modell wird zuerst mit dem blockierenden Reiz konditioniert, später wird dann gleichzeitig mit dem blockierenden und dem blockierten Reiz konditioniert, d.h zuerst Konditionierung mit CS_2 und später Konditionierung mit CS_2 und CS_3 .

Bei dem *Blocking* im Gallistel Modell handelt es sich um retroaktives (rückwirkendes) *Blocking* (Gallistel, 1990). Das Modell betrachtet nur die Gesamtzeiten T and die gesamten Zähler N . Es wird angenommen, dass die Auftretenshäufigkeit vom US, die einem CS zugeordnet wird, konstant ist. Sie ist somit unabhängig von der Reihenfolge der unterschiedlichen Intervalle.

4.4.3 Overshadowing (Überschattung von Reizen)

In einem *Overshadowing* Experiment überschattet der stärkere Reiz den weniger starken Reiz. Die beiden Reize treten nur gemeinsam auf. Das Versuchstier reagiert aber nur auf den stärkeren Reiz. Z.B Ton und Licht treten immer 5 Sekunden lang gemeinsam auf. Während der Darbietung von Ton und Licht erfolgt eine Verstärkung. Die Verstärkung tritt niemals ohne die beiden Reize Ton und Licht auf. Somit bekommt der Hintergrund als CS eine Rate von Null zugewiesen. Die Gleichungen für das Überschattungsexperiment sind dieselben wie für das *Blocking* Experiment. Dadurch, dass die Reize CS_2 und CS_3 immer gemeinsam auftreten, erhalten wir noch eine zusätzliche Bedingung zu den Bedingungen im Blocking Experiment (Gallistel, 1990).

$$T_2 = T_3 = T_{23} \quad (4.57)$$

Diese Bedingung führt dazu, dass die Determinante von T gleich Null wird, d.h wir erhalten unendlich viele Lösungen für unser Gleichungssystem. In unserem Beispiel mit dem Ton und dem Licht bedeutet dies, dass $\lambda_2 = \frac{1}{5}$ und $\lambda_3 = 0$ als Lösung gewählt werden könnte oder $\lambda_2 = 0$ und $\lambda_3 = \frac{1}{5}$. Eine andere Lösung wäre jedem CS die gleiche Rate zuzuordnen, nämlich $\lambda_2 = \lambda_3 = \frac{0.5}{5}$. Jede Kombination, die in ihrer Summe $\frac{1}{5}$ ergibt, ist eine mögliche Lösung, d.h. $\lambda_2 + \lambda_3 = \frac{1}{5}$. Dies wird von Gallistel auch das additive Prinzip genannt (Gallistel, 1990). Um eine eindeutige Lösung zu finden, wird das Prinzip der Prediktor Minimierung angewendet. Hierbei sollen redundante Reize eliminiert werden, um die Anzahl der Prediktoren zu minimieren. Es wird gefordert, dass die Lösung des RET Systems genommen wird, die die wenigsten Reize enthält. Dies sind die Alles oder Nichts Lösungen. In unserem Beispiel wird somit genau einem CS der US zugeordnet. Lösungen, die jedem CS einen Teil der US Auftretenshäufigkeit zuordnen, werden eliminiert.

Kapitel 5

Assoziieren mit den Lerntheorien - Anwendung der Lerntheorien auf das Erlernen von Wortassoziationen

5.1 Semantische Konditionierung

Eine frühe Anwendung der Theorie des bedingten Reflexes auf sprachliches Verhalten ist die semantische Konditionierung. Bei der semantischen Konditionierung werden Wörter als Reize verwendet. Ein Wort oder ein Satz wird als ein CS mit einem US verknüpft. Bei Konditionierungen höherer Ordnung werden Wörter miteinander verknüpft. Ein CS erster Ordnung bestimmt die Bedeutung des Wortes. Die Konditionierung erfolgt auf die Bedeutung des Wortes, nicht auf seinen Klang, seine Länge oder seine Buchstaben. Man spricht deshalb auch von der Konditionierung auf die Bedeutung des Wortes.

Eine Studie von Gregory Razran (Razran, 1939) belegt die Konditionierung auf die Bedeutung des Wortes. In seiner Studie verwendete Razran Nahrung als US (Lollipop, Sandwich) und maß den Speichelfluss bei College Studenten. Als CS leuchtete ein Wort auf, während die Teilnehmer am Essen waren. Wurde das Wort als CS ohne den US dargeboten, so speichelten die Teilnehmer. In einem weiteren Test wurden Synonyme und Homonyme Wörter zum Wort, das den CS darstellt, verwendet. Auch die synonymen und homonymen Wörter führten zu Speichelbildung. Die Reaktion Speichelbildung war bei den synonymen und homonymen Wörtern nicht so stark wie beim ursprünglichen Wort, das als CS diente.

Staats (Staats, 1968) hat gezeigt, dass die Bedeutungen von Wörtern

Ergebnisse von Konditionierungsprozessen sind. Zum Beispiel werden Wörter wie *Freude* und *Glück* häufig in angenehmen Stimulussituationen verwendet. Die Bedeutung der Wörter wird häufig durch die Paarung mit anderen Wörtern erworben. Staats gibt hierfür folgendes Beispiel: Das Kind hat häufig das Wort *verletzen* (hurt) in unangenehmen oder schmerzhaften Situationen gehört (z.B. ein Unfall, heiße Herdplatte). Das Wort *verletzen* wird mit aversiven Stimuli gepaart. Dadurch wird mit dem Wort *verletzen* eine negative emotionelle Bedeutung verbunden. Auf das Wort *verletzen* erfolgt dieselbe Reaktion wie auf den aversiven Reiz. Das Wort *verletzen* wird durch die Paarung mit einem aversiven Reiz zu einem konditionierten Stimulus. Die konditionierte emotionale Reaktion gibt dem Wort *verletzen* seine Bedeutung. Dieser Prozess wird nach Staats Konditionierung erster Ordnung auf die Bedeutung genannt (Staats, 1968).

US (ein aversiver Stimulus) \Rightarrow Reaktion auf aversiven Reiz
 Paarung von US (aversiver Stimulus) + CS Wort
 CS Wort \Rightarrow Reaktion auf aversiven Reiz

Wenn später dem Kind gesagt wird: *Dies ist gefährlich, du wirst dich verletzen.* (That is dangerous, you will get hurt.) Dann wird die emotionelle Reaktion auf das Wort *verletzen* (hurt) auf das Wort *gefährlich* (dangerous) übertragen. Diese Übertragung auf das Wort *gefährlich* ist eine Konditionierung höherer Ordnung, da der konditionierte Stimulus *verletzen* wie ein US wirkt.

Paarung von CS1 *verletzen* + CS2 *gefährlich*
 CS2 *gefährlich* \Rightarrow Reaktion auf aversiven Reiz

Wird dem Kind später mitgeteilt: *Motorräder sind gefährlich.* (Motorcycles are dangerous.) So wird die negative emotionale Reaktion auf das Wort *gefährlich* auf das Wort *Motorräder* übertragen. Der konditionierte Stimulus *gefährlich* wird zum US.

Paarung von CS2 *gefährlich* + CS3 *Motorräder*
 CS3 *Motorräder* \Rightarrow Reaktion auf aversiven Reiz

Das Kind wird auf alle drei Wörter, *Motorräder*, *gefährlich* und *verletzen*, mit der gleichen Reaktion wie auf den aversiven Reiz (US) reagieren.

Für das Erlernen von Wortassoziationen ist von Bedeutung, ob bei der Konditionierung höherer Ordnung eine Assoziation zwischen CS und US (S-S) gelernt wird oder zwischen CS und UR (S-R). Laut Schwartz & Reisberg

(Schwartz & Reisberg, 1991) wurden für beide Varianten, S-S und S-R, Beweise gefunden. Die Frage, ob eine S-S Assoziation oder eine S-R Assoziation bei der klassischen Konditionierung gelernt wird, bleibt somit ungelöst. Um das Erlernen von Wortassoziationen als eine Konditionierung höherer Ordnung darzustellen zu können, wird die erste Variante, das Bilden einer S-S Assoziation, benötigt. Dann werden im oberen Beispiel die drei Wörter *Motorräder*, *gefährlich* und *verletzen* miteinander assoziiert.

5.2 Operantes Lernen und Wortassoziationen

Allerdings kann die Produktion von Wörtern und von Wortassoziationen nicht allein durch die klassische Konditionierung erklärt werden, da nicht alle Wortbedeutungen auf einen unbedingten Reflexe zurückgeführt werden können. Nach Staats (Staats, 1968) spielt die operante Konditionierung eine wesentliche Rolle beim Erlernen von Wörtern und deren Bedeutung. Kinder erlernen Wörter, indem sie den Klang der Wörter der Eltern imitieren und danach von den Eltern verstärkt werden. Die Verstärkung bleibt aus, wenn der Klang der Kinder dem Wort der Eltern nicht ähnlich ist. Später lernen die Kinder Bilder und Aktivitäten (laufen, fahren usw.) mit Wörtern nach dem Prinzip der operanten Konditionierung zu verknüpfen. Zum Beispiel wird dem Kind das Bild eines Autos zusammen mit dem Wort *Auto* präsentiert. Wenn anschließend das Kind *Auto* sagt, wenn ein Auto zu sehen ist, so wird es durch das Verhalten der Eltern positiv verstärkt.

Im nächsten Schritt werden zwei Wörter hintereinander genannt. Z.B. *Auto fahren*, wenn ein Auto vorbeifährt. Das Kind bekommt von den Eltern als Stimuluswort das Wort *Auto* genannt und soll mit dem Wort *fahren* antworten. Antwortet das Kind mit dem gewünschten Wort, so wird diese Reaktion verstärkt. Das Kind lernt hierbei Wortpaare (*Auto, fahren*), es lernt auf das Stimuluswort *Auto* mit dem Responsewort *fahren* zu antworten.

Das Erlernen von Wortassoziationen kann als eine Konditionierung höherer Ordnung beschrieben werden. Bestimmte Wörter kommen häufig mit bestimmten anderen Wörtern in einem Text vor, z.B. *Brot* und *Butter*. Deswegen erwartet ein Leser, der ein Wort liest, gewisse andere Wörter in dessen Umgebung. Er hat gelernt, wenn er z.B. das Wort *Butter* liest, dann wird auch gleich das Wort *Brot* kommen. Wird seine Erwartung bestätigt, so erhöht dies seine Erwartung, dass beim nächsten Lesen von dem Wort *Butter* auch das Wort *Brot* erscheinen wird. Die Bestätigung der Erwartung wirkt wie eine Verstärkung bei der operanten Konditionierung. Wird seine Erwartung dagegen nicht bestätigt, so wird seine Erwartung abgeschwächt, dass beim nächsten Lesen von dem Wort *Butter* auch das Wort *Brot* erscheint.

Das Nichteintreten der Erwartung des Lesers wirkt wie das Ausbleiben der Verstärkung bei der operanten Konditionierung. Würde nach dem Wort *Butter* nie wieder das Wort *Brot* im Text folgen, so würde die Erwartung des Wortes *Brot* so abgeschwächt, dass der Leser, wenn er das Wort *Butter* liest, überhaupt nicht mehr das Wort *Brot* erwartet.

Dieser beschriebene Mechanismus des Erlernens von Assoziationen soll nun mit Hilfe von Computersimulationen überprüft werden. Einfache Lernvorgänge, wie Akquisition und Extinktion von Assoziationen, können mit Hilfe der Lerntheorien aus dem vorherigen Kapitel erklärt werden. Aus den drei Lerntheorien wurden drei Algorithmen zum Erlernen von Wortassoziationen abgeleitet, die in dem nächsten Abschnitt beschrieben werden.

5.3 Algorithmen für das Erlernen von Wortassoziationen

Für die Simulation wird ein Vokabular benötigt, das im nächsten Kapitel beschrieben wird. Alle drei Algorithmen, die für das Erlernen von Wortassoziationen verwendet werden sollen, verwenden die Fenstertechnik, die im 3. Kapitel beschrieben wurde.

Alle drei Algorithmen haben folgenden gemeinsamen Kern:

1. Zu Beginn wird eine konstante Anzahl von Wörtern eingelesen. Diese Wörter bilden das Wortfenster.
2. Es werden nur Wörter betrachtet, die in einem vorgegebenen Vokabular enthalten sind. Die anderen Wörter werden als sogenannte Nicht-Wörter markiert. Sie dienen als Platzhalter im Wortfenster.
3. Für alle Wortpaare im Wortfenster wird gemäß der drei Lerntheorien eine bestimmte Operation, die mit Lernen bezeichnet wird, durchgeführt. Die Wortpaare werden entweder nach der 1. Variante oder nach der 2. Variante aus dem Abschnitt Fenstertechnik im 3. Kapitel gebildet. Welche der beiden Varianten verwendet wird, hängt von der jeweiligen Lerntheorie ab.
4. Bei dem Estes Modell und dem Rescorla & Wagner Modell werden zusätzlich Wortpaare berücksichtigt, die aus einem Wort im Fenster und einem Wort aus dem Vokabular bestehen, das aber nicht im Wortfenster vorkommt. Auf diese Wortpaare wird eine Operation, die mit Interferenz oder Vergessen bezeichnet wird, angewendet. Wie diese Wortpaare zusammengestellt werden, hängt von der jeweiligen Theorie ab.

Das Gallistel Modell enthält als einziges Modell keine Regel für die Interferenz oder für Vergessen.

5. Nach Ausführung der obigen Operationen wird das Wortfenster um ein Wort nach rechts verschoben, d.h. es wird das nächste Wort aus dem Textkorpus eingelesen und rechts in das Wortfenster eingefügt. Das erste Wort im Fenster, das am linken Rand des Wortfensters steht, gleitet aus dem Textfenster.

Die Assoziationsstärken zwischen den Wörtern aus dem Vokabular werden in einer quadratischen Matrix abgespeichert. Für das Estes Modell ist diese Matrix symmetrisch, d.h. die Assoziationsstärke von einem Wortpaar (i, j) ist gleich der Assoziationsstärke des Wortpaares (j, i) . Die Reihenfolge der Wörter im Fenster wird beim Estes Modell nicht berücksichtigt. Für die Simulation des Rescorla & Wagner Modells und des Gallistel Modells wird eine nicht symmetrische Matrix benötigt, da bei diesen beiden Modellen die Reihenfolge der Wörter im Fenster berücksichtigt wird.

Um die Position eines Wortpaares in der Assoziationsmatrix festlegen zu können, wird jedem Wort aus dem Vokabular eine Zahl zwischen 0 und der Vokabulargröße minus 1 zugeordnet. Wenn ein Wort aus dem Textkorpus eingelesen wird, so wird nachgeschaut, ob das Wort im Vokabular vorhanden ist. Ist dies der Fall, so wird statt des Wortes seine zugehörige Zahl in das Wortfenster eingetragen. Kommt das Wort nicht im Vokabular vor, so wird eine -1 als Platzhalter für das Wort, das nicht im Vokabular vorkommt, in das Wortfenster eingetragen.

Mit Hilfe der Fenstertechnik soll das Lesen von Zeitungstexten simuliert werden. Die Wörter im Fenster repräsentieren dabei den Textausschnitt, der sich im Kurzzeitgedächtnis des Lesers befindet. Je nach Lerntheorie werden die Assoziationsstärken zwischen den Wörtern im Fenster unterschiedlich verstärkt.

In den folgenden Abschnitten werden die unterschiedlichen Algorithmen für das Erlernen von Wortassoziationen beschrieben.

5.3.1 Algorithmus für das Estes Modell

Nach der Stimulus Sampling Theorie werden zuerst die Merkmale aus der Umwelt wahrgenommen. Danach folgt eine Reaktion auf diese wahrgenommenen Elemente. Anschließend wird die Reaktion auf die wahrgenommenen Elemente durch ein Ereignis verstärkt oder nicht verstärkt. Wenn das verstärkende Ereignis eintritt, werden die wahrgenommenen Merkmale mit der Reaktion verknüpft. In der Regel werden beim Estes Modell zwei sich

ausschließende Ereignisse betrachtet, z.B. grünes Licht leuchtet oder rotes Licht leuchtet. Nach der Estes Theorie werden nicht die Merkmale mit dem Ereignis verknüpft, sondern die Merkmale werden mit der Reaktion auf diese verknüpft, wenn das verstärkende Ereignis eintritt. Deshalb muss überlegt werden, was die Reaktion auf die Wahrnehmung eines Wortes beim Erlernen von Wortassoziationen sein kann und was das verstärkende Ereignis ist. Eine mögliche Reaktion wäre die Erwartungshaltung des Lesers, wenn er ein Wort liest. Wenn der Leser ein Wort an einer bestimmten Stelle im Text liest, so erwartet er bestimmte Wörter in dessen Umgebung. Aufgrund des Wortes im Fenster sollen die anderen Wörter im Fenster vorausgesagt werden. Somit sind die Wörter im Fenster die situativen Merkmale. Zugleich sind sie aber auch Ereignisse, die vorhergesagt werden sollen. Eine richtige Voraussage führt zu einer Verstärkung der Voraussage. Deswegen muss zu allen Wörtern innerhalb des Fensters, die Lerngleichung aus dem Estes Modell angewendet werden, und für alle Wörter, die nicht im Fenster stehen, wird die Interferenzgleichung aus dem Estes Modell angewendet. Bei allen Wortpaaren, die nicht im Fenster vorkommen, ändert sich die Assoziationsstärke nicht. Es werden beim Erlernen von Wortassoziationen somit drei unterschiedliche Fälle betrachtet:

- Zwei Wörter i und j kommen gemeinsam im Fenster vor.
- Das Wort i steht im Fenster, aber nicht das Wort j , d.h. i tritt ohne j auf, oder das Wort j steht im Fenster und das Wort i tritt nicht auf, d.h. j tritt ohne i auf.
- Weder das Wort i noch das Wort j kommt im Fenster vor.

Der Leser eines Zeitungstextes, der simuliert werden soll, nimmt zu einem bestimmten Zeitpunkt nicht den ganzen Zeitungstext wahr, sondern nur einen gewissen Ausschnitt, der durch das Wortfenster festgelegt wird. Die wahrgenommenen Elemente sind nicht nur die Wörter im Wortfenster, sondern auch die Elemente aus der Umgebung, in der sich der Leser befindet. Für die Simulationen sind natürlich nur die Worte im Wortfenster als wahrgenommene Merkmale von Bedeutung. Während der Leser einen Satz liest, der durch das Fenster repräsentiert wird, erwartet der Leser mit einer bestimmten Wahrscheinlichkeit andere Wörter in dessen Umgebung. Er sieht ein Wort i aus dem Wortpaar (i, j) und erwartet, dass auch das Wort j auftritt. Wird seine Erwartung bestätigt, d.h. das Wort j ist auch im Fenster, so entspricht dies einer Verstärkung seiner Erwartung. Umgekehrt, wenn der Leser das Wort j aus dem Wortpaar (i, j) sieht, so erwartet er, dass auch das Wort i kommt. Kommt das Wort i dann tatsächlich, so wird seine Erwartung,

i kommt, bestätigt und führt zu einer Verstärkung seiner Erwartung, wenn das Wort j kommt, so kommt auch das Wort i . Diese Verstärkung führt zu einer Erhöhung der Assoziationsstärke nach der Lerngleichung aus dem Estes Modell.

$$a_{i,j}(n+1) = (1 - \theta) \cdot a_{i,j}(n) + \theta \quad (5.1)$$

Die Reihenfolge, in der das Wortpaar (i, j) auftritt, wird hier nicht berücksichtigt. Die beiden Möglichkeiten, wenn i auftritt dann tritt auch j auf, oder wenn j auftritt dann tritt auch i auf, werden in einem Ereignis zusammengefasst, i und j kommen gemeinsam im Fenster vor. Durch diese Zusammenfassung lässt sich der Speicherplatz, der für die Assoziationsmatrix benötigt wird, um die Hälfte reduzieren.

Bei Wettler & Rapp (Wettler, Rapp & Sedlmeier, 2005) werden die beiden zusammengefassten Ereignisse, wenn i auftritt dann tritt auch j auf, und wenn j auftritt dann tritt auch i auf, getrennt behandelt. Bei ihnen wird das Ereignis, wenn i auftritt dann tritt auch j auf, durch die bedingte Wahrscheinlichkeit $p(j|i)$ repräsentiert. Das Ereignis, wenn j auftritt dann tritt auch i auf, wird durch die bedingte Wahrscheinlichkeit $p(i|j)$ repräsentiert.

Bei Wettler & Rapp (Wettler, Rapp & Sedlmeier, 2005) wird nicht unterschieden, wie nah die Wortpaare im Fenster zusammenstehen, d.h es wird die 2. Variante der Wortpaarbildung aus dem Abschnitt Fenstertechnik im 3. Kapitel verwendet. Dagegen wird bei der Simulation des Estes Modells die 1. Variante der Wortpaarbildung verwendet, d.h. ein Wortpaar, das direkt nebeneinander im Fenster steht, wird bei einer Fensterlänge von k $(k-1)$ -mal verstärkt.

Ein Beispiel für ein Fenster der Länge 5 enthält die Wörter w_1 , w_2 , w_3 und w_4 . Zwischen dem Wort w_2 und w_3 steht noch ein Wort, das nicht im Vokabular vorkommt und mit einer -1 markiert ist.

$$\boxed{w_1 \ w_2 \ -1 \ w_3 \ w_4}$$

In dem aktuellen Lerndurchgang, der durch den obigen Fensterinhalt beschrieben wird, wird die Gleichung 5.1 auf die Wortpaare (w_1, w_2) , (w_1, w_3) , (w_1, w_4) , (w_2, w_3) , (w_2, w_4) und (w_3, w_4) angewendet.

Die Zelle $a_{i,j}$ in der Assoziationsmatrix enthält die Assoziationsstärke für das Wortpaar (w_i, w_j) . Der n -te Durchgang ist bei den Simulationen die n -te Fensterverschiebung. θ ist der Lernparameter aus dem Estes Modell. Die Bestimmung dieses Parameters wird im nächsten Kapitel im Abschnitt Parameterschätzung geschildert.

Im folgenden Abschnitt wird die Anwendung der Interferenzregel beschrieben. Diese wird angewendet, wenn nur ein Wort aus dem Wortpaar

im Fenster auftritt. Die Interferenzregel wird durch die Gleichung 5.2 beschrieben.

$$a_{i,j}(n+1) = (1 - \theta) \cdot a_{i,j}(n) \quad (5.2)$$

Durch die Interferenzregel wird die Assoziationsstärke des Wortpaares (i, j) abgeschwächt. Dies entspricht dem Fall, dass der Leser ein Wort aus dem Wortpaar sieht und dann das andere Wort aus dem Wortpaar erwartet. Seine Erwartung wird aber nicht bestätigt, somit bleibt die Verstärkung aus, d.h. die Assoziationsstärke wird nach der Gleichung 5.2 vermindert.

Würde beim obigen Beispiel das Vokabular aus den Wörtern w_1, w_2, w_3, w_4 und noch einem Wort w_5 bestehen, so würde die Assoziationsstärke von w_5 zu allen Wörtern im Fenster abgeschwächt, d.h. die Gleichung 5.2 wird für die Wortpaare $(w_1, w_5), (w_2, w_5), (w_3, w_5)$ und (w_4, w_5) ausgeführt.

Kommt keines der beiden Wörter aus dem Wortpaar (i, j) im n -ten Durchgang vor, so ändert sich die Assoziationsstärke des Wortpaares (i, j) nicht.

$$a_{i,j}(n+1) = a_{i,j}(n) \quad (5.3)$$

Wird im obigen Beispiel als nächstes das Wort w_5 aus dem Textkorpus eingelesen. So gleitet das Wort w_1 aus dem Fenster heraus, und das Wort w_5 wird hinten angehängt.

$$\boxed{w_2 \ -1 \ w_3 \ w_4 \ w_5}$$

Die Assoziationsstärken der Wortpaare $(w_2, w_3), (w_2, w_4), (w_2, w_5), (w_3, w_4), (w_3, w_5)$ und (w_4, w_5) werden nach der Gleichung 5.1 erhöht, während die Assoziationsstärken von Wort w_1 zu den Wörtern im Fenster abgeschwächt werden, da w_1 nicht mehr im Fenster steht. Bei Wortpaaren, die einen größeren Abstand zueinander haben, wird die Interferenzregel häufiger angewendet, als wenn die Wörter direkte Nachbarn sind.

Ein einfacher Algorithmus für die Berechnung von Wortassoziationen:

```

Initialisiere die gesamte Matrix a(i,j) mit 0
Wähle einen Wert zwischen 0 und 1 für T, den Lernparameter
Initialisiere FS, die Fensterlänge, mit einem geeignetem Wert
Lies das Vokabular ein
Initialisiere die Liste, die das Wortfenster repräsentiert

wiederhole
  lies ein Wort i aus dem Textkorpus ein
  wenn i im Vokabular enthalten ist

```

```

    füge i am Ende in das Wortfenster ein
  andernfalls
    füge ein -1 als Platzhalter in das Wortfenster ein
solange bis FS-1 Wörter vom Textkorpus eingelesen sind

wiederhole
  lies ein Wort i aus dem Textkorpus ein
  wenn i im Vokabular enthalten ist
    füge i am Ende in das Wortfenster ein
  andernfalls
    füge ein -1 als Platzhalter in das Wortfenster ein

Für alle Wortpaare (i < j),
  wenn die Worte i und j im Fenster stehen
    dann  $a(i,j) = (1-T) * a(i,j) + T$ 

    wenn i im Fenster ist, aber nicht j
      dann  $a(i,j) = (1-T) * a(i,j)$ 

    wenn j im Fenster ist, aber nicht i
      dann  $a(i,j) = (1-T) * a(i,j)$ 
  ende von für alle
  entferne das erste Wort im Wortfenster
solange bis das ganze Textkorpus komplett eingelesen worden ist

```

Ein Problem des obigen Algorithmus ist seine lange Laufzeit. Die zweite Schleife (wiederhole) wird für alle Wortpaare, wobei (i, j) und (j, i) das gleiche Wortpaar bezeichnen, aus dem Vokabular abgearbeitet. Die Schleife (Für alle Wortpaare) muss bei einer Vokabulargröße von m für $\frac{m(m-1)}{2}$ - Wortpaare durchlaufen werden. Die Laufzeit dieser Schleife lässt sich verbessern, indem nur die Wortpaare im Fenster neu berechnet werden und nicht alle Wortpaare. Dann ist die Laufzeit der Schleife (Für alle Wortpaare) nicht mehr von der Vokabulargröße abhängig, sondern von der Fenstergröße, die sehr viel kleiner als die Vokabulargröße ist. Damit nur die Wortpaare im Fenster neu berechnet werden müssen, muss zu jedem Wortpaar zusätzlich abgespeichert werden, wann es zum letzten Mal berechnet wurde und wie oft die beiden Wörter aus dem Wortpaar seit der letzten Berechnung alleine im Fenster vorgekommen sind. Zu jedem Wortpaar (i, j) werden folgende Informationen benötigt:

1. Einen Zähler z_i für jedes Wort i aus dem Vokabular. Der Wert von z_i

gibt an, wie häufig das Wort i bisher im Fenster vorgekommen ist. Tritt das Wort i im Fenster auf, so wird z_i um eins erhöht.

2. Die aktuelle Assoziationsstärke $a_{i,j}$ seit der letzten Berechnung. Die letzte Berechnung erfolgte in dem Durchgang, in dem das Wortpaar (i, j) gemeinsam im Fenster auftrat.
3. Wann ein Wortpaar zum letzten Mal berechnet wurde. Dies wird in der Variable $h_{i,j}$ gespeichert. Sie enthält die Summe der beiden Zähler z_i und z_j zum Zeitpunkt der letzten Berechnung. Die Variable $h_{i,j}$ erhält nur dann einen neuen Wert, wenn auch die Assoziationsstärke $a_{i,j}$ neu berechnet wird.

Zu Beginn der Simulation wird die aktuelle Assoziationsstärke, die Variable $h_{i,j}$ und der Zähler für jedes Wort mit Null initialisiert.

$$a_{i,j} = 0 \quad (5.4)$$

$$h_{i,j} = 0 \quad (5.5)$$

$$z_i = 0 \quad (5.6)$$

Die Assoziationsstärke zu einem Wortpaar (i, j) soll nur dann neu berechnet werden, wenn das Wortpaar (i, j) gemeinsam im Textfenster vorkommt. Steht nur ein Wort aus dem Wortpaar im Textfenster, nennen wir dies i , so wird nur der Zähler z_i für dieses Wort i erhöht. Tritt nun das Wortpaar (i, j) in einem späteren Durchgang im Fenster zusammen auf, so muss zuerst bestimmt werden, wie häufig Gleichung 5.2 zur Berechnung der Assoziationsstärke verwendet werden muss. Dies lässt sich wie folgt aus den Zählern von i und j und der Variable $h_{i,j}$ ablesen:

$$k = z_i + z_j - h_{i,j} \quad (5.7)$$

Bei der Gleichung 5.7 bezeichnet z_i den Zähler für das Wort i und z_j den Zähler für das Wort j . In z_i steht, wie häufig das Wort i bisher im Fenster vorgekommen ist. Entsprechend gibt z_j an, wie häufig das Wort j bisher im Fenster vorgekommen ist. In $h_{i,j}$ steht der Wert von $z_i + z_j$ zum Zeitpunkt der letzten Berechnung der Assoziationsstärke $a_{i,j}$. Wurde die Assoziationsstärke $a_{i,j}$ im n -ten Durchgang, d.h. nach der n -ten Fensterverschiebung, neu berechnet, so enthält die Variable $h_{i,j}$ die Summe der beiden Zähler z_i und z_j im n -ten Durchgang. Muss im $(n + k + 1)$ -ten Durchgang die Assoziationsstärke $a_{i,j}$ neu berechnet werden, so können sich die beiden Zähler z_i und z_j seit

dem n -ten Durchgang verändert haben. Dieser Fall ist eingetreten, wenn i ohne j oder j ohne i im Fenster aufgetreten ist. Der Wert von $h_{i,j}$ ist seit dem n -ten Durchgang nicht verändert worden, da die Assoziationsstärke $a_{i,j}$ noch nicht neu berechnet worden ist. Der Wert von k gibt an, wie häufig seit der letzten Berechnung, die war im n -ten Durchgang, nur ein Wort aus dem Wortpaar (i, j) im Fenster vorgekommen ist.

- Ist $z_i + z_j = h_{i,j}$, dann ist weder i noch j seit der letzten Berechnung der Assoziationsstärke von i und j alleine im Fenster vorgekommen.
- Ist $z_i + z_j > h_{i,j}$, so ist k -mal nur ein Wort aus dem Wortpaar (i, j) im Fenster vorgekommen.

Deswegen wird die Gleichung 5.2 k -mal hintereinander ausgeführt, bevor die Assoziationsstärke von i und j neu berechnet werden kann. Dies lässt sich mit folgender Gleichung in einem Schritt durchführen:

$$a_{i,j}(n+k) = (1-\theta)^k \cdot a_{i,j}(n) \quad (5.8)$$

In zwei Fällen wird die obige Gleichung nicht auf ein Wortpaar (i, j) angewendet:

1. Das Wortpaar (i, j) ist bisher noch nicht gemeinsam im Textfenster vorgekommen, d.h. die Assoziationsstärke des Wortpaares ist gleich Null.
2. Im letzten Durchgang war das Wortpaar (i, j) auch schon zusammen im Textfenster, d.h. k hat den Wert Null.

Nach Anwendung von Gleichung 5.8 auf das Wortpaar (i, j) wird die Gleichung 5.1 angewendet, und wir erhalten die Assoziationsstärke für das Wortpaar (i, j) im $(n+k+1)$ -ten Durchgang. Nach der Berechnung der Assoziationsstärke von dem Wortpaar (i, j) erhält noch die Variable $h_{i,j}$ die Summe der beiden Zähler z_i und z_j .

$$h_{i,j} = z_i + z_j \quad (5.9)$$

Die Variable $h_{i,j}$ enthält jetzt die Summe von z_i und z_j im $(n+k+1)$ -ten Durchgang.

Durch diesen Algorithmus brauchen nur die Wortpaare im Fenster neu berechnet werden.

Nach dem das gesamte Textkorpus eingelesen worden ist, muss noch einmal für alle Wortpaare bestimmt werden, wie oft die Gleichung 5.8 für jedes

Wortpaar ausgeführt werden muss. Dies kann mit Hilfe der Gleichung 5.7 bestimmt werden. Ist k in der Gleichung 5.7 für ein Wortpaar größer als Null, so muss für dieses Wortpaar die Gleichung 5.8 ausgeführt werden, da seit der letzten Berechnung k -mal ein Wort aus diesem Wortpaar alleine im Fenster vorgekommen ist.

Nach der Berechnung der Assoziationsstärken können die Assoziationen zu einem Stimuluswort bestimmt werden. Um die Assoziationen zu einem vorgegebenen Stimuluswort zu erhalten, wird zunächst die zugehörige Spalte in der Assoziationsmatrix bestimmt. In dieser Spalte stehen die Assoziationsstärken zu dem gegebenen Stimuluswort. Diese Spalte kann als Assoziationsvektor bezeichnet werden. Dieser enthält in den Zeilen die Assoziationsstärken zu den anderen Wörtern aus dem Vokabular. Die Primärantwort wird aus diesem Assoziationsvektor bestimmt, indem das Wort mit der maximalen Assoziationsstärke ermittelt wird.

5.3.2 Algorithmus für das Rescorla & Wagner Modell

Im Rescorla & Wagner Modell wird zwischen konditionierten und unkonditionierten Stimuli unterschieden. Um das Rescorla & Wagner Modell simulieren zu können, muss deshalb festgelegt werden, was der CS und was der US beim Erlernen von Wortassoziationen sein soll. Beim Lesen eines Textes erwartet der Leser, dass nach dem Lesen eines Wortes, das als ein CS gesehen werden kann, ein bestimmtes Wort, das man als US betrachten könnte, folgt. Alle Wörter aus dem Vokabular können sowohl CS als auch US sein. Durch das Wortfenster wird festgelegt, welche Wörter als CS und welches Wort als US zusammen auftreten. Das letzte Wort im Fenster wird als US definiert, da beim klassischen Konditionieren üblicherweise der US auf den CS folgt. In einem Lerndurchgang tritt genau ein US auf. Dies ist das letzte Wort im Fenster. Die Wörter vor dem US im jeweiligen Fenster repräsentieren die einzelnen Reize, die im aktuellen Durchgang aktiv sind. Bei einer Fensterlänge von n ist das Wort an der n -ten Position im Fenster der US, während die Wörter an den Positionen 1 bis $n-1$ die Reize CS_1 bis CS_{n-1} darstellen. Für ein Wort i , das einen dargebotenen Reiz repräsentiert, wird die Veränderung der Assoziationsstärke zu dem Wort j im Fenster, das den aufgetretenen US darstellt, nach folgender Gleichung berechnet:

$$\Delta a_{i,j}(m) = \alpha_i \beta_j (\lambda - \sum_k a_{k,j}(m-1)) \quad (5.10)$$

Die Summe wird über alle $n-1$ Worte im Fenster gebildet, d. h. über alle Assoziationsstärken von dem Wort, das den US darstellt, zu den Wörtern im Fenster, die die Reize darstellen. In die Berechnung der Assoziationsstärke

geht nicht nur die vorherige Assoziationsstärke ein, sondern auch noch die Assoziationsstärken der anderen Wörter im Fenster zu dem letzten Wort im Fenster. Befinden wir uns gerade im m -ten Durchgang, so wird die Assoziationsstärke von Wort i zu dem Wort j folgendermaßen bestimmt.

$$a_{i,j}(m) = a_{i,j}(m - 1) + \Delta a_{i,j}(m) \quad (5.11)$$

Die Assoziationsstärke zwischen dem Wort i und dem Wort j wird durch die obige Gleichung verstärkt, d.h. es wird gelernt. Die obigen beiden Formeln werden auf alle dargebotenen Reize, d.h. alle $n-1$ Wörter im Fenster, angewendet.

Im folgenden Abschnitt wird die Anwendung von Gleichung 5.11 auf ein Fenster mit 5 Wörtern demonstriert:

$$\boxed{w_1 \ w_2 \ -1 \ w_3 \ \boxed{w_4}}$$

Das Fenster enthält die Wörter w_1, w_2, w_3 und w_4 . Das mit -1 gekennzeichnete Element im Fenster steht für ein Wort im Fenster, das nicht im Vokabular vorkommt. Das letzte Wort im Fenster repräsentiert den aktuellen US. Dies ist das Wort w_4 . Die Wörter w_1, w_2 und w_3 repräsentieren die aktiven Reize. Die Assoziationsstärken von den aktiven Reizen zum aktuellen US werden mit Hilfe der Gleichung 5.11 erhöht. Dies sind die Assoziationsstärken der Wortpaare (w_1, w_4) (w_2, w_4) und (w_3, w_4) . Dabei hängt die Veränderung der Assoziationsstärke $\Delta a_{i,j}$ von der Summe der Assoziationsstärken der Wörter im Fenster zu dem Wort w_4 ab. Die Veränderung der Assoziationsstärken wird für das obige Beispiel folgendermaßen berechnet:

$$\Delta a_{i,4} = \alpha_i \beta_4 (\lambda - (a_{1,4} + a_{2,4} + a_{3,4})) \quad (5.12)$$

i nimmt in der obigen Gleichung die Werte 1,2 und 3 an. Die neuen Assoziationsstärken $a_{i,4}(neu)$ für die Wortpaare (w_1, w_4) , (w_2, w_4) und (w_3, w_4) ergeben sich dann folgendermaßen:

$$a_{i,4}(neu) = a_{i,4} + \Delta a_{i,4} \quad (5.13)$$

Alle Wörter aus dem Vokabular, die nicht im Fenster stehen, werden als ein US betrachtet, der im aktuellen Durchgang nicht aufgetreten ist. Die Assoziationsstärken von den Wörtern im Fenster zu den Wörtern im Vokabular, die nicht im aktuellen Fenster stehen, werden deswegen vermindert, d. h. ($\lambda = 0$) für diese Wörter.

$$\Delta a_{i,l}(m) = \alpha_i \beta_l (0 - \sum_k a_{k,l}(m - 1)) \quad (5.14)$$

Die Änderung der Assoziationsstärke von dem Wort i im Fenster zu einem Wort l , das nicht im Fenster steht, berechnet sich aus der vorherigen Assoziationsstärke zu dem Wort l minus der Summe aller Assoziationsstärken von den $n-1$ Wörtern im Fenster zu dem Wort l . Es wird nur die Assoziationsstärke von den $n-1$ Wörtern im Fenster neu berechnet, da das letzte Wort als US betrachtet wird und somit kein CS ist. Die Assoziationsstärke von dem Wort i im Fenster zu einem Wort l wird dann folgendermaßen bestimmt:

$$a_{i,l}(m) = a_{i,l}(m - 1) - \Delta a_{i,l}(m) \quad (5.15)$$

Wegen $\lambda = 0$ wird $\Delta a_{i,l}$ negativ, die Assoziationsstärke zwischen dem Wort i und dem Wort l wird abgeschwächt, d.h. es wird vergessen. Die Abschwächung der Assoziationsstärken soll an folgendem Fensterinhalt demonstriert werden:

$$\boxed{w_1 \ w_2 \ -1 \ w_3 \ \boxed{w_4}}$$

Wäre im obigen Beispiel noch das Wort w_5 im Vokabular, so müssen die Assoziationsstärken von den aktuellen Reizen zu dem Wort w_5 abgeschwächt werden. Denn das Wort w_5 steht für einen US, der im aktuellen Durchgang nicht aufgetreten ist. Die aktuellen Reize sind die Wörter w_1 , w_2 und w_3 . Das Wort w_4 im Fenster ist kein Reiz, da es den aktuellen US repräsentiert. Die Assoziationsstärken der Wörter, die die Reize repräsentieren, zu dem Wort w_5 werden vermindert. Dies sind die Wortpaare (w_1, w_5) , (w_2, w_5) und (w_3, w_5) . Dabei ist die Abschwächung der Assoziationsstärken, wie beim Lernen, von der Summe der Assoziationsstärken zu dem Wort w_5 abhängig. Die Veränderung der Assoziationsstärke zum Wort w_5 wird folgendermaßen berechnet:

$$\Delta a_{i,5} = -\alpha_i \beta_5 (a_{1,5} + a_{2,5} + a_{3,5}) \quad (5.16)$$

Die neuen Assoziationsstärken $a_{i,5}(neu)$ der Wortpaare (w_1, w_5) , (w_2, w_5) und (w_3, w_5) werden dann folgendermaßen aus den bisherigen Assoziationsstärken berechnet:

$$a_{i,5}(neu) = a_{i,5} + \Delta a_{i,5} \quad (5.17)$$

Die Assoziationsstärke von einem Wortpaar kann auch negativ werden. Dies passiert genau dann, wenn $\Delta a_{i,j}$ größer als die Assoziationsstärke des Wortpaares ist.

Die Assoziationsstärken aller Wortpaare aus dem Vokabular werden in einer Matrix abgespeichert, die im Gegensatz zum Estes Modell, nicht symmetrisch ist. Um die Assoziationen zu einem Wort zu erhalten, kann beim

Estes Modell sowohl die Spalte als auch die Zeile zum Auslesen der Assoziationen verwendet werden. Beim Rescorla & Wagner Modell wird zur Bestimmung der Assoziationen zu einem Wort j die j -te Zeile der Assoziationsmatrix benötigt, da in der Zeile der Assoziationsmatrix der CS und in der Spalte der US steht. Um die Assoziationsstärke eines CS zu einem US zu überprüfen, wird in einem Konditionierungsexperiment der CS gegeben und es wird geschaut, ob eine bedingte Reaktion erfolgt und wie stark diese Reaktion ist. Bei den Wortassoziationen werden die Stimuluswörter als Reize betrachtet. Für ein Stimuluswort j wird in der Assoziationsmatrix in der j -te Zeile das Wort mit maximaler Assoziationsstärke ermittelt, das anschließend als Primärantwort von dem Computerprogramm ausgegeben wird. Die Assoziationsstärke der Primärantwort zu dem Stimuluswort j entspricht dabei der Reaktionsstärke auf den CS.

Die Assoziationsstärken können beim Rescorla & Wagner Modell im Gegensatz zum Estes Modell auch negativ werden. Da für die Simulationen $\lambda = 1$ gewählt wird, können die Assoziationsstärken zwischen 1 und -1 variieren.

Es folgt ein einfacher Algorithmus für die Berechnung von Wortassoziationen mit dem Rescorla & Wagner Modell:

Initialisierung wie beim Estes Modell

Wähle für alpha und beta Werte zwischen 0 und 1

Einlesen der FS-1 Wörter in das Wortfenster
(wie beim Estes Modell)

wiederhole

 einlesen eines Wortes aus dem Textkorpus und
 einfügen in das Wortfenster erfolgt wie beim
 Estes Modell

 us = das letzte Wort in dem Wortfenster

 delta = 0

 sum = 0

 Für alle cs in dem Wortfenster

 wenn cs ungleich us

 dann sum = sum + a(cs,us)

 Für alle cs in dem Wortfenster

 wenn cs ungleich us

 dann $a(cs,us) = a(cs,us) + \alpha * \beta * (1 - \text{sum})$

```

Für alle Wörter u aus dem Vokabular, die nicht im Fenster
enthalten sind
  sum = 0
  Für alle cs in dem Wortfenster
    wenn cs ungleich u und cs ungleich us
      dann sum = sum + a(cs,u)
  Für alle cs in dem Wortfenster
    wenn cs ungleich u und cs ungleich us
      dann a(cs,u) = a(cs,u) - alpha * beta * sum
ende für alle
entferne das erste Wort aus dem Wortfenster
solange bis das ganze Textkorpus komplett eingelesen worden ist

```

5.3.3 Algorithmus für das Gallistel Modell

Auch beim Gallistel Modell wird zwischen dem CS und dem US unterschieden. Wie beim Rescorla & Wagner Modell werden die $k-1$ Wörter im Fenster als Reize betrachtet, während das letzte Wort den US repräsentiert. Die Zeit im Gallistel Modell wird so interpretiert, dass ein Fenster genau einer Zeiteinheit darstellt.

Für jedes Wort wird ermittelt, wie häufig es als CS im Fenster vorkommt. Dies entspricht der kumulativen Darbietungsdauer eines Reizes im Gallistel Modell. Diese wird in der Variable t_i für ein Wort i abgespeichert. In der Matrix $t_{i,j}$ steht das gemeinsame Vorkommen des Wortes i mit dem Wort j im Fenster als CS. Dabei werden nur die Wörter an den Positionen 1 bis $k-1$ im Fenster betrachtet, da nur diese Reize darstellen. Das k -te oder das Wort an der letzten Position im Fenster ist der US. Für diesen wird die Anzahl des gemeinsamen Auftretens mit den aktiven Reizen, also allen Wörtern an den Positionen von 1 bis $k-1$ im Fenster, in der Matrix n abgespeichert. Im folgenden Beispiel sind die Reize durch die Wörter w_1 , w_2 und w_3 repräsentiert:

w_1	w_2	-1	w_3	w_4
-------	-------	------	-------	-------

Das letzte Wort im Fenster ist der aktuelle US. Die Reize w_1 , w_2 und w_3 werden dargeboten, wenn der US w_4 auftritt. Dies führt zu einer Inkrementierung von den Zellen $n_{1,4}$, $n_{2,4}$ und $n_{3,4}$ in der Matrix n , da in der Matrix n die Auftretenshäufigkeit eines US mit einem CS vermerkt wird.

In $t_{i,j}$ wird die kumulative gemeinsame Darbietungsdauer zweier Reize i und j gespeichert. Dies ist bei den Wörtern die gemeinsame Auftretenshäufigkeit als CS im Fenster. In dem obigen Beispiel sind das die Wortpaare (w_1, w_2) , (w_1, w_3) und (w_2, w_3) , d.h. die Zellen $t_{1,2}$, $t_{1,3}$ und $t_{2,3}$ in der

Matrix t müssen um 1 erhöht werden. Anschließend werden noch die Zähler für die Reize w_1 , w_2 und w_3 um 1 erhöht, dies sind t_1 , t_2 und t_3 , da diese Reize eine Zeiteinheit dargeboten wurden.

Nachdem das Fenster über den gesamten Textkorpus geschoben worden ist, wird die Zeitkoeffizientenmatrix T berechnet. Diese erhalten wir, indem wir für jede Zeile i die Auftretenshäufigkeiten $t_{i,j}$ durch die Häufigkeiten t_i des Wortes i teilen. Die gemeinsame Auftretenshäufigkeit eines Wortpaares als CS wird durch die Einzelhäufigkeit eines Wortes als CS dividiert.

$$T_{i,j} = \frac{t_{i,j}}{t_i} \quad \forall i, j \quad (5.18)$$

Nach dieser Berechnung steht in $T_{i,j}$ das Verhältnis, Anzahl der Kookkurrenzen von Wort i als CS und Wort j als CS zu der Gesamthäufigkeit des Wortes i als CS.

Im nächsten Schritt wird für alle Zeilen aus der Matrix n das Verhältnis von Anzahl der Kookkurrenzen von Wort i als CS und Wort j als US zu der Gesamthäufigkeit des Wortes i als CS gebildet.

$$N_{i,j} = \frac{n_{i,j}}{t_i} \quad \forall i, j \quad (5.19)$$

Eine Spalte in der Matrix $N_{i,j}$ enthält die unkorrigierten Auftretensraten von dem US_j mit den jeweiligen Reizen CS_i . Die unkorrigierten Auftretensraten für einen US mit den jeweiligen Reizen stehen im Gallistel Modell in dem Vektor λ_u .

Um die korrekten Raten zu erhalten, wird die inverse Matrix von der Matrix T gebildet und anschließend mit der Matrix N multipliziert. Die korrekten Auftretensraten für einen US stehen dann in den Spalten der Matrix L . Im Gallistel Modell stehen diese in den Vektoren $(\lambda_1, \dots, \lambda_m)$. Somit enthält eine Spalte von L die Vektoren $(\lambda_1, \dots, \lambda_m)$ für einen US, wobei mit m die Größe des Vokabulars bezeichnet wird.

$$L = T^{-1} \cdot N \quad (5.20)$$

Die Ermittlung der Wortassoziationen aus der Matrix L erfolgt genauso wie beim Rescorla & Wagner Modell.

Für die Berechnung von Wortassoziationen mit dem Gallistel Modell werden folgende Variablen benötigt:

- Matrix $T(i,j)$, sie entspricht der Matrix T aus dem Gallistel Modell
- Matrix $N(i,j)$, sie enthält für jeden US den Vektor λ_u aus dem Gallistel Modell. Das Wort j repräsentiert den US, während Wort i den Reiz darstellt.

- Matrix $L(i,j)$, sie enthält die einzelnen Vektoren $(\lambda_1, \dots, \lambda_m)$ aus dem Gallistel Modell
- Vektor $tc(i)$, er enthält die Auftretenshäufigkeit des Wortes i als CS im Fenster

Der Algorithmus für die Berechnung der Wortassoziationen:

```
Initialisiere Matrix  $T(i,j)$  und Matrix  $N(i,j)$  mit 0
Initialisiere Vektor  $tc(i)$  mit 0
Initialisierung von FS
Vokabular einlesen und Initialisierung des
Wortfensters erfolgt wie beim Estes Modell
```

wiederhole

```
  Einlesen der ersten FS-1 Wörter vom Textkorpus
  in das Fenster erfolgt wie beim Estes Modell
```

wiederhole

```
  einlesen eines Wortes aus dem Textkorpus und
  einfügen in das Wortfenster erfolgt wie beim
  Estes Modell
```

```
  us = letztes Wort in dem Wortfenster
```

```
  Für alle cs in dem Wortfenster
```

```
    wenn cs ungleich us
```

```
    dann erhöhe  $N(cs,us)$  um 1
```

```
    erhöhe  $tc(cs)$  um 1
```

```
  ende für alle cs
```

```
  Für alle cs1 in dem Wortfenster
```

```
    Für alle cs2, die hinter dem cs1 im Wortfenster stehen
```

```
      wenn (cs1 ungleich us) und (cs2 ungleich us)
```

```
      dann erhöhe  $T(cs1,cs2)$  um 1 und
```

```
        erhöhe  $T(cs2,cs1)$  um 1
```

```
    ende für alle cs2
```

```
  ende für alle cs1
```

```
  entferne das erste Wort aus dem Fenster
```

```
solange bis das ganze Textkorpus komplett eingelesen worden ist
```

```

Für alle Worte i im Vokabular
  Für alle Worte j im Vokabular
    wenn i gleich j
      dann T(i,i)=1
    andernfalls
      wenn tc(i) ungleich 0
        dann
          T(i,j) = T(i,j) / tc(i)
          N(i,j) = N(i,j) / tc(i)
        ende dann
      ende für alle j
    ende für alle i

```

T = Berechne die inverse Matrix von T

L = Multipliziere die Matrix T mit der Matrix N

In der Matrix N wird zuerst die gemeinsame Auftretenshäufigkeit von CS und US im Fenster abgespeichert. Dies wird in der ersten Schleife (für alle cs) für alle Wörter, die einen Reiz in dem Fenster darstellen, durchgeführt. Zusätzlich wird in dieser Schleife noch die Auftretenshäufigkeit eines Wortes als CS, $tc(cs)$, im Fenster gezählt. Die Matrix T enthält die gemeinsame Auftretenshäufigkeit zweier Reize $cs1$ und $cs2$ im Fenster.

Die gemeinsame Auftretenshäufigkeit zweier Reize im aktuellen Fenster wird in der ersten geschachtelten Schleife (Für alle $cs1$ im Wortfenster) gezählt. Nach dem Einlesen des gesamten Textkorpus enthält die Matrix T die gemeinsamen Auftretenshäufigkeiten der Reize im Fenster, und die Matrix N enthält die gemeinsamen Auftretenshäufigkeiten von CS und US im Fenster.

In der zweiten geschachtelten Schleife (Für alle Wörter i im Vokabular) werden die Elemente in der Diagonale der Matrix T auf 1 gesetzt, da die Auftretenshäufigkeit eines Wortes mit sich selbst dividiert durch seine Auftretenshäufigkeit 1 ergibt ($T(i,i) = tc(i)$). Die Diagonale der Matrix N besteht nur aus Nullen, da ein Wort nicht gleichzeitig ein CS und ein US sein kann. Die Berechnung der neuen Matrix T erfolgt nach Gleichung 5.18 und die Berechnung der neuen Matrix N erfolgt nach Gleichung 5.19.

Zum Schluss wird die Matrix T invertiert. Für die Invertierung der Matrix T wird die MTL (Matrix Template Library) verwendet. Die MTL ist im Internet frei verfügbar und kann unter der Adresse <http://www.osl.iu.edu/research/mtl/> heruntergeladen werden. Anschließend wird die invertierte Matrix T mit der Matrix N multipliziert. Für die Matrixmultiplikation wird ebenfalls die MTL verwendet. Nach der Matrixmultipli-

kation enthält die Matrix L die korrekten Auftretensraten. Die Spalten der Matrix L enthalten die einzelne Vektoren $(\lambda_1, \dots, \lambda_m)$ aus dem Gallistel Modell. Bei den im Kapitel 6 beschriebenen Simulationen des Gallistel Modells konnte die inverse Matrix immer bestimmt werden, d.h. die Matrix konnte in allen durchgeführten Simulationen bestimmt werden.

Für das Erlernen von Wortassoziationen entspricht die Umsetzung des Gallistel Modells der Umsetzung des Rescorla & Wagner Modell im Equilibrium. Danks untersucht in (Danks, 2003), wann sich das Rescorla & Wagner Modell im Equilibrium befindet. Nach Danks befindet sich das Rescorla & Wagner Modell im Equilibrium unter der Bedingung, dass $\lambda = 1$ gewählt wird und dass für Vergessen und Erlernen die gleichen Parameter gewählt werden, wenn folgendes Gleichungssystem erfüllt ist:

$$\begin{pmatrix} 1 & P(C_2|C_1) & \dots & P(C_n|C_1) \\ P(C_1|C_2) & 1 & \dots & P(C_n|C_2) \\ \vdots & & & \\ \vdots & & & \\ P(C_1, C_n) & P(C_2|C_n) & \dots & 1 \end{pmatrix} \cdot \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ \vdots \\ V_n \end{pmatrix} = \begin{pmatrix} P(US|C_1) \\ P(US|C_2) \\ \vdots \\ \vdots \\ P(US|C_n) \end{pmatrix} \quad (5.21)$$

C_i steht für den CS mit der Nummer i . V_i ist die gesuchte Assoziationsstärke zwischen dem CS C_i und dem US. Der Vektor V entspricht dem Vektor λ_c aus dem Gallistel Modell. $P(C_j|C_i)$ ist die bedingte Wahrscheinlichkeit dafür, dass bei gegebenen Reiz C_i der Reiz C_j auftritt. Beim Erlernen von Wortassoziationen ist die bedingte Wahrscheinlichkeit $P(C_j|C_i)$ durch die gemeinsame Auftretenshäufigkeit $h(i, j)$ von zwei Wörtern i und j als CS dividiert durch die Auftretenshäufigkeit $h(i)$ von j als CS gegeben.

$$P(C_j|C_i) = \frac{h(i, j)}{h(i)} \quad (5.22)$$

Im Gallistel Modell ist $\frac{h(i, j)}{h(i)} = \frac{t_{i, j}}{t_i}$ (siehe Gleichung 5.18). In $t_{i, j}$ steht die gemeinsame Auftretenshäufigkeit zweier Wörter, die beide einen CS im Fenster repräsentieren. t_i ist die gesamte Auftretenshäufigkeit von einem Wort als CS im Fenster. t_i ist nicht die Häufigkeit des Wortes im gesamten Textkorpus, sondern die Auftretenshäufigkeit des Wortes im Fenster als CS. Diese ist abhängig von der Fensterlänge, während die Worthäufigkeit im Textkorpus unabhängig von der Fensterlänge ist. Somit entspricht die Matrix mit den bedingten Wahrscheinlichkeiten in 5.21 der Matrix T im Gallistel Modell für das Erlernen von Wortassoziationen.

Der Vektor $P(US|C_i)$ mit den bedingten Wahrscheinlichkeiten, dass bei gegebenen CS der US auftritt, entspricht einer Spalte in der Matrix N. Denn

in der Matrix $N_{i,j}$ steht die Auftretenshäufigkeit von dem Wort i als CS mit dem Wort j als US dividiert durch die Auftretenshäufigkeit des Wortes i als CS (siehe Gleichung 5.19).

$$P(US|C_i) = \frac{n_{i,US}}{t_i} \quad (5.23)$$

Somit entspricht die Umsetzung des Gallistel Modells für das Erlernen von Wortassoziationen dem Rescorla & Wagner Modell im Equilibrium, wenn beim Rescorla & Wagner Modell der Parameter λ gleich 1 gewählt wird, und wenn für Lernen und Vergessen die gleichen Parameter gewählt werden.

Kapitel 6

Simulationen

Im vorherigen Kapitel wurden Algorithmen zum Erlernen von Wortassoziationen beschrieben. In diesem Kapitel werden die Simulationen, die mit diesen Algorithmen durchgeführt wurden, beschrieben.

Zunächst wird im nächsten Abschnitt das Lernmaterial, die Textkorpora, vorgestellt und deren Aufbereitung für die Simulationen.

Aus Laufzeit- und Speicherplatzgründen konnten nicht alle Wörter, die in den Zeitungskorpora vorkamen, für die Berechnung der Wortassoziationen verwendet werden. Deshalb wurde ein Vokabular definiert, das alle 100 Stimulus Wörter aus den Assoziationsnormen und die zugehörigen Responsewörter, die von mindestens 5 Teilnehmern genannt wurden, enthielt. Dies waren insgesamt 814 Wörter. Von diesen 814 Wörtern kamen 802 Wörter in den verwendeten Textkorpora vor.

Für den Vergleich der Lernmodelle untereinander wurde ein Gütemaß für die Bewertung der berechneten Assoziationen benötigt. Auf das verwendete Gütemaß, das aus vier Kriterien besteht, wird im Abschnitt Kriterien eingegangen.

Die drei Lernmodelle haben alle freie Parameter, die experimentell ermittelt werden mussten. Ein freier Parameter, den alle Simulationen gemeinsam haben, ist die Größe des Wortfensters (Fensterlänge). Dazu kommen noch die freien Parameter, die die einzelnen Lernmodelle enthalten. Das Gallistel Modell ist das einzige von den drei Lernmodellen, das keinen freien Parameter enthält. Das Estes Modell enthält als freien Parameter die Lernrate θ . Das Rescorla & Wagner Modell enthält drei freie Parameter, die Reizstärke des CS, die Reizstärke des US und die maximale Assoziationsstärke. Die Bestimmung dieser freien Parameter wird im Abschnitt Parameterschätzung erläutert.

6.1 Korpora

Für die Berechnung von Wortassoziationen werden große maschinenlesbare Textsammlungen benötigt. Als Lernmaterial für die Simulation des Erlernens von Wortassoziationen wurde eine auf CD-ROM gespeicherte Sammlung aller Texte der Süddeutschen Zeitung der Jahrgänge 1995 bis 1998 verwendet. Dadurch stand eine große Textmenge zur Berechnung von Wortassoziationen zur Verfügung. Diese Textsammlung besteht nicht nur aus politischen Texten, denn die Süddeutsche Zeitung enthält auch einen großen Feuilletonteil.

Die vier Jahrgänge der Süddeutschen Zeitung wurden nach der alten Rechtschreibung verfasst. Im nachfolgenden Abschnitt und in den folgenden Kapiteln wird mit dem Textkorpus SZ95 der Jahrgang 1995 der Süddeutschen Zeitung bezeichnet, der Jahrgang 1996 der Süddeutschen Zeitung wird mit SZ96, der Jahrgang 1997 der Süddeutschen Zeitung wird mit SZ97, der Jahrgang 1998 der Süddeutschen Zeitung wird mit SZ98 und mit SZ werden alle vier Jahrgänge zusammen bezeichnet. Das Textkorpus SZ ist somit ein Textkorpus, das die Textkorpora SZ95, SZ96, SZ97 und SZ98 in der angegebenen Reihenfolge enthält.

Die Textkorpora SZ95 und SZ96 sind ungefähr gleich groß, sind aber insgesamt kleiner als das Textkorpus SZ97 und das Textkorpus SZ98. Die Textkorpora SZ95 und SZ96 enthalten ca. 29 Millionen Wörter, während die SZ97 ca. 1 Millionen mehr Wörter als das Textkorpus SZ95 und das Textkorpus SZ96 enthält. Das Textkorpus SZ98 enthält ca. 34 Millionen Wörter. Das Textkorpus SZ kommt somit auf ca. 123 Millionen Wörter.

Ein Textkorpus von den vier Textkorpora sollte als Kalibrierkorpus verwendet werden, um die freien Parameter der einzelnen Lernmodelle zu bestimmen. Als Kalibrierkorpus wurde das Textkorpus SZ98 ausgewählt, da es das größte Textkorpus von den vier Textkorpora SZ95, SZ96, SZ97 und SZ98 war. Das Textkorpus SZ98 wird in diesem Kapitel im Abschnitt Parameterschätzung zur Bestimmung der freien Parameter verwendet.

Die anderen Textkorpora SZ95, SZ96 und SZ97 wurden als Testkorpora verwendet, um zu Testen, ob die freien Parameter, die mit dem Kalibrierkorpus ermittelt wurden, auch bei anderen Textkorpora gute Ergebnisse erzielen würden. Das Textkorpus SZ ist ein besonderes Textkorpus, da es die anderen Textkorpora enthält. Es sollte aber trotzdem auch als Testkorpus verwendet werden, da hierdurch ein sehr großes Textkorpus für die Assoziationsberechnungen zur Verfügung stand. Die Simulationen, die mit den Textkorpora durchgeführt wurden, werden im nächsten Kapitel beschrieben.

6.2 Lemmatisierung

Bevor die Simulationen durchgeführt werden konnten, wurden die Textkorpora lemmatisiert, d.h. alle flektierten Wortformen wurden durch ihre Grundformen ersetzt. Das Wort *gegangen* wurde z.B. auf die Grundform *gehen*, das Wort *Häuser* wurde auf die Grundform *Haus* zurückgeführt. Durch die Lemmatisierung wurde die Anzahl verschiedener Wörter reduziert und die absoluten Worthäufigkeiten der Wörter in der Grundform erhöht. Wie stark die Worthäufigkeit durch die Lemmatisierung erhöht wurde, hing stark von den Wörtern selber ab. Bei einigen mehrdeutigen Wörtern, die auf eine Grundform zurückgeführt wurden, führte die Lemmatisierung zu einer starken Zunahme der Worthäufigkeit. Der Wortstamm *sein* hatte zum Beispiel nach der Lemmatisierung des Textkorporus SZ98 eine Worthäufigkeit von 714186. Vor der Lemmatisierung kam das Wort *sein* nur 56766 mal in dem Textkorporus SZ98 vor. Dieser starke Anstieg der Worthäufigkeit kam von den beiden Bedeutungen von *sein*. Die Grundform *sein* war einmal durch die Adjektive *seine*, *seiner* usw. oder durch die Verben *ist* und *bin* entstanden. Da alle flektierten Formen von der Grundform *sein* sehr häufig im Textkorporus vorkamen, führte dies zu so einer hohen Worthäufigkeit des Wortes *sein*. Bei anderen häufigen Wörtern, wie z.B. *Mann* oder *Frau*, wurde die Worthäufigkeit durch die Lemmatisierung nur verdoppelt. Die Erhöhung der Worthäufigkeiten durch die Lemmatisierung hing stark von den Gebrauch der flektierten Formen ab. Denn die Worthäufigkeit des Wortes *Polizei* wurde durch die Lemmatisierung nur wenig erhöht, von 10169 in dem nicht lemmatisierten Textkorporus SZ98 auf 10377 Wörter in dem lemmatisierten Textkorporus SZ98. Dies lag daran, dass die flektierten Formen von dem Wort *Polizei* in dem Textkorporus SZ98 wesentlich seltener als die Grundform verwendet wurden. Dagegen wurden die flektierten Formen von dem Verb *wollen* viel häufiger verwendet als die Grundform. Dies führte zu einer starken Erhöhung der Worthäufigkeit von dem Verb *wollen*, von 17007 auf 69818, durch die Lemmatisierung. Im Mittel lag die Worthäufigkeit der Wörter aus dem Vokabular bei 1306, wenn das nicht lemmatisierte Textkorporus SZ98 für die Bestimmung der Worthäufigkeiten verwendet wurde. Nach der Lemmatisierung des Textkorporus SZ98 erhöhte sich die mittlere Worthäufigkeit der Wörter aus dem Vokabular von 1306 auf 3124. Die mittlere Worthäufigkeit wurde als Mittelwert berechnet.

Für die Lemmatisierung wurde eine Ersetzungstabelle mit MORPHY (Lezius, Rapp & Wettler, 1998) generiert. MORPHY ist ein von Lezius entwickeltes Programm für die morphologische Analyse und Synthese und die kontextsensitive Wortartenbestimmung der deutschen Sprache. Das Programmpaket MORPHY besteht aus zwei Komponenten, ein Morphologiesystem und eine

Taggingkomponente. Das Morphologiesystem ermittelt alle möglichen grammatikalischen Beschreibungen eines Wortes unabhängig von seinem Kontext. Hierfür wird ein Lexikon mit 324.000 Wortformen, die auf einem Lexikon mit 50.000 Stammformen basieren, verwendet. Es enthält alle Wörter aus dem Wörterbuch *Wahrig Deutsches Wörterbuch* (Wahrig, 1997). Das Morphologiesystem generiert zu einem gegebenen Wort zuerst eine Liste aller möglichen Stammformen. Die Stammformen entstehen durch das Abschneiden aller möglichen Präfixe, Infixe und Endungen, versetzen von ß und ss und anderer Umlaute (z.B. Haus - Häuser). Nachdem alle möglichen Stammformen des gegebenen Wortes ermittelt wurden, wird jede Stammform im Lexikon gesucht. Falls sie im Lexikon enthalten ist, werden alle möglichen Wortformen zu dieser Stammform generiert. Anschließend werden die Stammformen gestrichen, aus denen sich nicht das Originalwort herleiten lässt. Nach der morphologischen Analyse erhält man zu einem gegebenen Wort eine Liste mit den zugehörigen Stammformen und den zugehörigen Tags. Die Tags beinhalten die grammatikalischen Informationen zu einem gegebenen Wort. Da bei der morphologischen Analyse die Wörter ohne ihren Kontext betrachtet werden, können Mehrdeutigkeiten nicht aufgelöst werden. Für die Auflösung von Mehrdeutigkeiten ist die 2. Komponente von MORPHY, der Tagger, der auch den Kontext eines Wortes berücksichtigt, zuständig. Der Tagger verwendet die bedingten Wahrscheinlichkeiten von Tagsequenzen. MORPHY verwendet den Trigram - Algorithmus von Church für das prohabilitische Tagging (Church, 1988). Der Lemmatisierer im MORPHY Programm verwendet die Ausgabe des Taggers, um mehrdeutige Wortformen mit mehr als einer Stammform aufzulösen.

Um die Textkorpora für die Assoziationsberechnungen mit MORPHY zu lemmatisieren, müsste vor der eigentlichen Lemmatisierung das Textkorpus am besten zuerst getaggt werden. Dies war aber aus Rechenzeitgründen nicht möglich, da MORPHY pro Sekunde ca. 20 Wörter taggen kann (Lezius, Rapp & Wettler, 1996) und die vier Jahre SZ aus ca. 120 Mio. Wörter bestehen. Der Rechner würde zum Taggen der SZ ca. 96 Tage benötigen. Deswegen wurde für die Lemmatisierung der Textkorpora nur die 1. Komponente von MORPHY verwendet. Mit Hilfe der morphologischen Analyse wurde eine Ersetzungstabelle generiert, die zu jedem deutschen Wort seine Grundform und die zugehörigen flektierten Formen enthielt. Insgesamt enthielt die mit MORPHY generierte Ersetzungstabelle ca. 78.000 Wortgrundformen. Die Ersetzungstabelle bildet den Kern des Lemmatisierers des ATA2-Programms (Böhnisch, 1999) mit dem die vier Jahrgänge Süddeutsche Zeitung lemmatisiert worden sind. Dieser Lemmatisierer kann zwar keine Mehrdeutigkeiten auflösen, benötigt dafür aber wesentlich weniger Rechenzeit. Dafür muss aber in Kauf genommen werden, dass Mehrdeutigkeiten falsch aufgelöst werden.

Ein Beispiel für eine falsche Auflösung von Mehrdeutigkeiten ist das Wort *weiß*, das als Stimuluswort in den Assoziationsnormen vorkommt. Mit dem Wort *weiß* kann einmal das Adjektiv *weiß* (z.B. weiße Wand) gemeint sein, oder das Verb *wissen* (z.B. er weiß). Der Lemmatisierer aus dem ATA2-Programm kann die Mehrdeutigkeit von dem Wort *weiß* nicht auflösen. Das Wort *weiß* wurde von dem Lemmatisierer immer durch das Wort *wissen* ersetzt. Durch die falsche Auflösung von Mehrdeutigkeiten entstand eine fehlerhafte Lemmatisierung. Somit konnten die Assoziationsberechnungen zu dem Stimuluswort *weiß* ebenfalls nicht korrekt sein.

Der ATA2 - Lemmatisierer ersetzt alle Satzzeichen, Bindestriche, Klammern und Anführungsstriche durch Punkte. Wörter, die nicht in der Ersetzungstabelle des Lemmatisierers enthalten sind, werden bei der Lemmatisierung unverändert übernommen. Die Zahlen werden vom Lemmatisierer in einzelne Ziffern, die jeweils durch Leerzeichen getrennt sind, zerlegt.

6.3 Kriterien

Nach den Assoziationsberechnungen mit Hilfe der in Kapitel 5 beschriebenen Algorithmen für die einzelnen Lernmodelle erhält man als Ergebnis eine Assoziationsmatrix. Die Assoziationsmatrix enthält in den Zeilen die Stimuluswörter und in den Spalten die möglichen Responsewörter. Jedes Wort aus dem Vokabular kann sowohl ein Stimuluswort als auch ein Responsewort sein. Für den Vergleich der unterschiedlichen Lernmodelle untereinander wurden aber nur die Stimuluswörter aus den Assoziationsnormen verwendet.

Die Wortassoziationen zu einem gegebenen Stimuluswort werden folgendermaßen ermittelt:

- Extrahiere den zum Stimuluswort gehörenden Zeilenvektor aus der Assoziationsmatrix. Der Zeilenvektor enthält die Assoziationsstärken aller möglichen Responsewörter zu einem gegebenen Stimuluswort.
- Sortiere die zugehörigen Responsewörter in dem Zeilenvektor nach der Assoziationsstärke. Durch diese Sortierung erhält man eine nach Assoziationsstärken sortierte Liste mit den zugehörigen Responsewörtern. Diese sortierte Liste wird als Rangreihe bezeichnet.

Für alle 100 Stimuluswörter aus den Assoziationsnormen wird eine Rangreihe, wie oben beschrieben, aufgestellt. In dieser Rangreihe, die durch die Sortierung nach der Assoziationsstärke entsteht, steht das Responsewort mit der größten Assoziationsstärke zum Stimuluswort an erster Stelle. Dieses Responsewort ist die vorausgesagte Primärantwort, sie steht auf dem ersten

Rangplatz in der Rangreihe. Das Responsewort mit der zweitgrößten Assoziationsstärke zum Stimuluswort steht auf dem zweiten Rangplatz in der Rangreihe, usw. . An der letzten Stelle in der Rangreihe steht das Responsewort mit der geringsten Assoziationsstärke zum Stimuluswort.

Um zu überprüfen, wie gut die berechneten Wortassoziationen mit den von Russell und Jenkins erhobenen Wortassoziationen übereinstimmen, wird ein Gütemaß benötigt. Dieses Gütemaß besteht aus den folgenden vier Kriterien:

1. Kriterium: Die Anzahl der Primärantworten der Teilnehmer, die vom Programm richtig vorausgesagt werden.
2. Kriterium: Es werden alle Antworten, die von mindestens 5 Teilnehmern für ein Stimuluswort genannt wurden, mit Ausnahme der Primärantwort berücksichtigt. Es wird die Anzahl dieser Antworten ermittelt, die bei der Berechnung auf den ersten 10 Rangplätzen stehen. Die Primärantwort wird dabei nicht mitgezählt. Dies ist die Anzahl richtig vorausgesagter Responsewörter auf den Rangplätzen 1 bis 10. Anschließend wird für jedes Stimuluswort die Anzahl richtig vorausgesagter Responsewörter aufsummiert.
3. Kriterium: Die Anzahl der Primärantworten der Teilnehmer, die vom Programm auf Rangplatz 1 oder 2 vorausgesagt werden.
4. Kriterium: Die Anzahl der Primärantworten der Teilnehmer, die vom Programm auf den ersten fünf Rangplätzen vorausgesagt werden.

Bei dem ersten Kriterium muss für jedes der 100 Stimuluswörter aus dem Assoziationsversuch ermittelt werden, ob die Primärantwort der Teilnehmer auch von dem Programm vorausgesagt wird, d.h. die Primärantwort steht bei den berechneten Assoziationen auf Rangplatz 1. Stimmen für ein Stimuluswort die Primärantwort der Teilnehmer mit dem berechneten Responsewort auf Rangplatz 1 überein, so erhöht sich die Anzahl richtig vorausgesagter Primärantworten um 1. Für das erste Kriterium muss somit die Anzahl der Stimuluswörter ermittelt werden, bei denen die Primärantwort der Teilnehmer mit der berechneten Primärantwort übereinstimmt.

Für das zweite Kriterium werden die ersten zehn berechneten Responsewörter, dies sind die Responsewörter auf den Rangplätzen 1 bis 10, zu dem jeweiligen Stimuluswort untersucht. Von diesen zehn Responsewörtern werden die Responsewörter gezählt, die von mindestens 5 Teilnehmern zu dem jeweiligen Stimuluswort genannt wurden. Die Primärantwort wird bei diesem Kriterium nicht mitgezählt.

Es folgt ein Beispiel, an dem ersichtlich wird, welche Responsewörter beim zweiten Kriterium gezählt werden. In der Tabelle 6.1 sind die zehn stärksten berechneten Responsewörter zu dem Stimuluswort *essen* aufgeführt.

Rangplatz	Assoziationsstärke	Responsewort
1	0.0002979	trinken
2	0.0001131	Mahlzeit
3	7.262e-05	Brot
4	6.288e-05	Fleisch
5	6.286e-05	Suppe
6	6.176e-05	Schokolade
7	5.920e-05	Salat
8	5.904e-05	Blumenkohl
9	5.763e-05	Reis
10	5.56e-05	schlafen

Tabelle 6.1: Berechnete Assoziationen zu dem Stimuluswort *essen*

In der Tabelle 6.2 sind die Responsewörter der Teilnehmer zum Stimuluswort *essen* aufgeführt.

Die Responsewörter *trinken*, *Mahlzeit*, *schlafen*, *Fleisch* und *Brot* der Teilnehmer werden auch vom Programm auf den ersten zehn Rangplätzen vorausgesagt. Das Responsewort *trinken* ist die Primärantwort, d.h. sie wurde von den meisten Teilnehmern im Assoziationsversuch genannt. Bei den berechneten Assoziationen zu dem Stimuluswort *essen* wird ebenfalls *trinken* als Primärantwort vorausgesagt, d.h. sie steht auf Rangplatz 1 bei den berechneten Assoziationen. Für das zweite Kriterium wird die Primärantwort der Teilnehmer nicht mehr berücksichtigt, da sie schon bei den anderen drei Kriterien berücksichtigt wird. Für das Stimuluswort *essen* fallen somit die Responsewörter *Mahlzeit*, *schlafen*, *Fleisch* und *Brot* der Teilnehmer auf die ersten zehn Rangplätze. Das Wort *Brot* z.B. wurde von 7 Teilnehmern im Assoziationsversuch genannt. Bei den vom Programm berechneten Assoziationen steht das Wort *Brot* auf Rangplatz 3. Insgesamt befinden sich in dem obigen Beispiel vier Responsewörter der Teilnehmer auf den ersten zehn Rangplätzen. Die Anzahl der richtig vorausgesagten Responsewörter (ohne Berücksichtigung der Primärantwort) für das Stimuluswort *essen* ist somit vier. Für jedes Stimuluswort wird die Anzahl der richtig vorausgesagten Responsewörter der Teilnehmer auf den ersten zehn Rangplätzen, wie gerade beschrieben, ermittelt. Anschließend wird die Anzahl der richtig vorausgesagten Responsewörter für jedes Stimuluswort aufsummiert. Diese Summe gibt an, wie viele Responsewörter der Teilnehmer insgesamt vom Programm auf

Responsewort	Anzahl
trinken	76
Hunger	40
gut	23
Mahlzeit	10
schlafen	9
Speise	8
Zeit	8
Brot	7
Teller	7
Fleisch	6
Mittagessen	5
Nahrung	5
Tisch	5

Tabelle 6.2: Assoziationen der Teilnehmer zu dem Stimuluswort *essen*. In der rechten Spalte sind die Anzahl der Teilnehmer aufgeführt, die das jeweilige Responsewort im Assoziationsversuch genannt hatten. In der linken Spalte sind Responsewörter aufgeführt, die von mindestens fünf Teilnehmern im Assoziationsversuch genannt wurden.

den ersten zehn Rangplätzen ausgegeben werden. Die Primärantwort wird dabei nicht als Responsewort betrachtet. Aus der Anzahl der richtig vorausgesagten Responsewörter kann ermittelt werden, wie viele Responsewörter der Teilnehmer außer der Primärantwort im Durchschnitt für ein Stimuluswort auf den ersten zehn Rangplätzen zu finden sind.

Für die Bewertung der berechneten Assoziationen wird zuerst die Anzahl richtig berechneter Primärantworten berücksichtigt (Kriterium 1). Je mehr Primärantworten richtig vorausgesagt werden, je besser ist das Ergebnis in Bezug auf dieses Kriterium. Die anderen Maße werden gemäß ihrer Reihenfolge gewichtet. Auch bei den anderen Kriterien erhält man eine Zahl. Je größer diese Zahl ist, umso besser ist das Ergebnis in Bezug auf das jeweilige Kriterium. Existieren mehrere Ergebnisse, die durch die vier Kriterien nicht weiter unterschieden werden können, so wird das Ergebnis, das bei einer kürzeren Fensterlänge erzielt wurde, als das bessere Ergebnis gewertet.

Die Anzahl richtig vorausgesagter Primärantworten wird als primäres Maß, d.h. das erste Kriterium wurde stärker gewichtet als die anderen Kriterien, verwendet, da die Primärantwort bei vielen Stimuluswörtern von deutlich mehr Teilnehmern genannt wurde als die anderen Responsewörter. Bei dem Stimuluswort *Mann* wurde die Primärantwort *Frau* von 171 Teilnehmern

genannt, während das nächste Responsewort *Weib* nur noch von 15 Personen genannt wurde. Bei 26 Stimuluswörtern wurde die Primärantwort von mehr als 100 Teilnehmern genannt. Nicht immer waren sich die Teilnehmer bei der Primärantwort einig. Dies zeigt sich zum Beispiel bei dem Stimuluswort *Arzt*. Die Primärantwort *Krankheit* wurde von nur 38 Teilnehmern genannt, während 37 Teilnehmer zu diesem Stimuluswort das Wort *Doktor* als Responsewort nannten. Die Stimuluswörter *Hammelfleisch* und *Hammer* hatten sogar zwei Primärantworten, da die beiden häufigsten Responsewörter von der gleichen Anzahl Teilnehmer genannt wurden.

Bei den meisten Stimuluswörtern (59) wurde jedoch die Primärantwort von 20 Teilnehmern mehr genannt als das zweithäufigste Responsewort.

Während das erste Kriterium nur die Primärantwort berücksichtigt, berücksichtigt das zweite Kriterium alle anderen Responsewörter zu den jeweiligen Stimuluswörtern, die von mindestens 5 Teilnehmern genannt wurden. Für jedes Stimuluswort wird gezählt, wie viele von seinen Responsewörtern (ohne die Primärantwort) bei den Simulationen auf den ersten zehn Rangplätzen sind. Anschließend wird die Anzahl der Responsewörter auf den ersten zehn Rangplätzen für jedes Stimuluswort aufsummiert. Bei dem zweiten Kriterium wird nicht berücksichtigt, wie häufig die Responsewörter zu einem Stimuluswort genannt wurden, sie müssen nur von mindestens 5 Teilnehmern genannt worden sein. Durch dieses Kriterium soll erreicht werden, dass möglichst viele Responsewörter der Teilnehmer auch vom Programm auf den ersten zehn Rangplätzen ausgegeben werden.

Das dritte Kriterium berücksichtigt die Anzahl der richtig vorausgesagten Primärantworten auf den ersten und zweiten Rangplatz. Während das vierte Kriterium die Anzahl der richtig vorausgesagten Primärantworten auf den ersten fünf Rangplätzen angibt. Somit erfüllen die Stimuluswörter, die das erste Kriterium erfüllen, auch das dritte Kriterium, und die Stimuluswörter, die das dritte Kriterium erfüllen, erfüllen auch das vierte Kriterium.

Im nächsten Abschnitt wird beschrieben, wie die vier Kriterien zur Parameterschätzung verwendet wurden.

6.4 Parameterschätzung

Alle drei Modelle haben die Fensterlänge als einen freien Parameter, der experimentell bestimmt werden musste. Bei dem Estes Modell kam noch der Lernparameter θ als freier Parameter hinzu. Das Rescorla & Wagner Modell hatte zusätzlich zur Fensterlänge noch drei freie Parameter, α für die Reizstärke des CS, β für die Reizstärke des US und λ die maximale Assoziationsstärke, die bestimmt werden mussten.

Alle Simulationsreihen wurden mit einer Fensterlänge von 12 gestartet. Die Fensterlänge wurde für die folgenden Simulationen iterativ um 1 erhöht, solange bis entweder eine Fensterlänge von 100 erreicht wurde oder die Ergebnisse in Bezug auf die vier Kriterien bei mehreren Fensterlängen sich kontinuierlich verschlechterten, so dass kein besseres Ergebnis mehr zu erwarten war.

Wenn möglich, sollte für alle drei Lernmodelle eine gemeinsame Fensterlänge gefunden werden, bei der alle Lernmodelle gute Ergebnisse in Bezug auf die vier Kriterien erzielten.

Es war zu erwarten, dass bei den Simulationen die Worthäufigkeit eine entscheidende Rolle spielt. Im nächsten Abschnitt wird beschrieben, wie die Worthäufigkeit in den einzelnen Lernmodellen berücksichtigt wurde.

6.4.1 Worthäufigkeitseffekt

Mit Worthäufigkeitseffekt bezeichnet man den Effekt, dass sehr häufige Wörter häufig als Responsewort berechnet werden, obwohl die Teilnehmer im Assoziationsversuch sehr häufige Wörter eher selten als Responsewort nennen. Der Worthäufigkeitseffekt wurde bereits in den Arbeiten von Wettler & Rapp beobachtet (siehe Kapitel 3).

Die Worthäufigkeiten der fünf häufigsten Wörter aus dem Vokabular im Kalibrierkorpus sind:

- sein, mit einer Worthäufigkeit von 714.186
- Seite, mit einer Worthäufigkeit von 110.466
- wollen, mit einer Worthäufigkeit von 69.818
- gut, mit einer Worthäufigkeit von 50.908
- gehen, mit einer Worthäufigkeit von 43.183

Die hohe Worthäufigkeit des Wortes *sein* resultiert aus der Mehrdeutigkeit des Wortes und daraus, dass beide Bedeutungen sehr häufige Wörter im Textkorpus sind (siehe hierzu den Abschnitt Lemmatisierung). Das Wort *Seite* kommt so häufig in dem Textkorpus SZ98 vor, da jede Seite in der Zeitung oben rechts das Wort *Seite* mit der zugehörigen Seitenzahl enthält. Für die Worthäufigkeiten der anderen drei sehr häufigen Wörter sind keine besonderen Gründe ersichtlich.

Um den Worthäufigkeitseffekt bei der Berechnung der Wortassoziationen zu verringern, wurden folgende zwei Lösungsansätze verwendet:

1. Nach der Berechnung werden die berechneten Assoziationen zu einem gegebenen Stimuluswort mit einem Faktor gewichtet, der von der Worthäufigkeit des Responsewortes abhängig ist. Responsewörter mit niedrigen Worthäufigkeiten erhalten einen größeren Faktor als Wörter mit höheren Worthäufigkeiten.
2. Die Lernparameter und die Parameter für Vergessen werden für die einzelnen Wortpaare unterschiedlich gewählt. Wortpaare mit höheren Worthäufigkeiten erhalten einen kleineren Wert als Wortpaare mit niedrigen Worthäufigkeiten.

Für die Gewichtung der berechneten Assoziationen im ersten Lösungsansatz wurde ein Maß aus der Informationstheorie verwendet. Der Mathematiker Claude Shannon hat eine Formel entwickelt mit dessen Hilfe der Informationsgehalt eines Zeichens bestimmt werden kann. Nach Shannon ist der Informationsgehalt eines Zeichens um so größer, desto seltener das Zeichen im Text vorkommt. Überträgt man dieses auf Wörter, so bedeutet dies: Je häufiger ein Wort im Textkorpus vorkommt, je geringer ist sein Informationsgehalt. Der Informationsgehalt eines Wortes im Textkorpus ist dann folgendermaßen definiert:

$$I_j = -\log_2 p(j) \quad (6.1)$$

Mit $p(j)$ wird die Auftretenswahrscheinlichkeit eines Wortes j im Textkorpus bezeichnet. Sie berechnet sich aus der Worthäufigkeit des Wortes j dividiert durch die Anzahl der Wörter im Textkorpus.

Je größer die Auftretenswahrscheinlichkeit eines Wortes j ist, desto geringer ist sein Informationsgehalt. Seltene Wörter bekommen einen höheren Informationsgehalt als häufige Wörter zugewiesen.

Beim ersten Lösungsansatz zur Verringerung des Worthäufigkeitseffekts werden die Assoziationen zu einem gegebenen Stimuluswort nach der Berechnung mit dem Informationsgehalt des jeweiligen Responsewortes gewichtet. Die Gewichtung erfolgt durch eine Multiplikation der berechneten Assoziationsstärke mit dem Informationsgehalt des Responsewortes.

$$g_{i,j} = a_{i,j} \cdot I_j \quad (6.2)$$

In der Gleichung 6.2 wird mit i das Stimuluswort und mit j das Responsewort bezeichnet. Mit $g_{i,j}$ wird die gewichtete Assoziationsstärke bezeichnet, während mit $a_{i,j}$ die berechnete Assoziationsstärke bezeichnet wird.

Beim zweiten Lösungsansatz zur Verringerung des Worthäufigkeitseffekts, werden die Parameter, Lernparameter und der Parameter für Vergessen,

im Lernmodell für die einzelnen Wortpaare variiert. Für die Variation der Parameter wird der Informationsgehalt verwendet. Wortpaare mit hohen Worthäufigkeiten sollen einen kleineren Parameter als Wortpaare mit geringen Worthäufigkeiten erhalten. Dass Wortpaare mit häufigen Wörtern einen kleineren Parameter bekommen als Wortpaare, die nicht so häufig vorkommen, könnte mit folgendem Resultat, das bei Untersuchungen der Augenbewegung von Lesern gewonnen wurde, gerechtfertigt werden (Rayner & Polatsek, 1989):

Eine Untersuchung der Augenbewegung von Lesern zeigt, dass seltene und unerwartete Wörter länger vom Leser fixiert werden als häufige vorher-sagbare Wörter. Wörter werden oft gar nicht fixiert, wenn sie häufig, kurz und vorhersagbar sind.

Der Lernparameter θ beim Estes Modell gibt an, wie viele Elemente aus der Umwelt wahrgenommen werden. Wird ein häufiges Wort kürzer fixiert als ein selteneres Wort, so könnte dies bedeuten, dass es weniger stark vom Leser wahrgenommen wird als das selteneres Wort. Der Informationsgehalt eines Wortes dient in den Simulationen des Estes Modells als Maß dafür, wie stark ein Wort vom Leser wahrgenommen wird. Beim Rescorla & Wagner Modell gibt der Informationsgehalt eines Wortes die Reizstärke des Wortes an.

Der Informationsgehalt der Wörter aus dem Vokabular liegt in einem Wertebereich von 5 bis 25. Das häufigste Wort *sein* hat einen Informationsgehalt von 5.58 und die seltensten Wörter (Wörter mit einer Worthäufigkeit von 1) haben einen Informationsgehalt von 25.02 in dem Textkorpus SZ98. Der Lernparameter θ muss aber im Wertebereich von 0 und 1 liegen. Um dies zu erreichen, wird der Informationsgehalt der Wörter im Wortpaar so mit einem Skalierungsfaktor multipliziert, dass der Lernparameter einen Wert zwischen 0 und 1 erhält.

$$\theta_{i,j} = F \cdot (I_i \cdot I_j) \quad (6.3)$$

Der Skalierungsfaktor F erhält für alle Wortpaare denselben Wert. Deswegen erhalten Wortpaare mit häufigen Wörtern durch die Gleichung 6.3 einen kleineren Wert als Wortpaare mit niedrigen Worthäufigkeiten. Denn häufige Wörter haben einen kleineren Informationsgehalt als seltene Wörter. Durch die Gleichung 6.3 erhalten Wortpaare mit niedrigen Worthäufigkeiten eine steilere Lernkurve als Wortpaare mit häufigen Wörtern.

6.4.2 Hill-Climbing Algorithmus

Um die freien Parameter in den einzelnen Lernmodellen zu ermitteln, wurde der Hill-Climbing Algorithmus verwendet. Der Hill-Climbing Algorithmus arbeitet nach dem Bergsteigerprinzip: Gehe niemals bergab.

Der Hill-Climbing Algorithmus benötigt einen Startpunkt, eine Suchmenge S und eine Bewertungsfunktion f . Die Bewertungsfunktion gibt an, welcher Punkt aus der Menge S das beste Ergebnis liefert. Der Punkt, bei dem die Bewertungsfunktion maximal wird, ist der gesuchte Punkt.

Die Umgebung von einem Punkt oder Wert wird folgendermaßen definiert:

$$N(x) = \{y \in S : \|x - y\| \leq \epsilon\} \quad \epsilon > 0 \quad (6.4)$$

Durch ϵ kann die Umgebung von x verkleinert oder vergrößert werden.

Von einem gegebenen Startpunkt x aus, wird die Suche nach einer besseren Lösung in Bezug auf die Bewertungsfunktion gestartet. Aus der Umgebung $N(x)$ wird eine konstante Anzahl von Punkten y ausgewählt. Für diese Punkte wird die Bewertungsfunktion $f(y)$ berechnet. Es wird der Punkt ermittelt, für den die Bewertungsfunktion maximal wird. Dieser Punkt wird mit y_m bezeichnet. Ist $f(y_m) > f(x)$, so ist ein besseres Ergebnis gefunden worden, und y_m ist der neue Startpunkt, indessen Umgebung nach noch besseren Ergebnissen gesucht werden soll. Hat der Punkt y_m kein besseres Ergebnis geliefert, d.h. $f(y_m) \leq f(x)$, so terminiert der Algorithmus.

Bei dem iterativen Hill-Climbing Algorithmus können folgende Probleme auftreten (Michalewicz & Fogel, 2000):

- Der Hill-Climbing Algorithmus findet eventuell nur ein lokales Optimum, d.h. das globale Optimum wird nicht gefunden.
- Generell ist es nicht möglich eine obere Schranke für die Laufzeit des Algorithmus anzugeben.

Um den ersten Punkt zu entschärfen, werden in (Michalewicz & Fogel, 2000) unterschiedliche Startwerte für den Hill-Climbing Algorithmus empfohlen.

Als Bewertungsfunktion für die Parameterschätzung wurden die vier Kriterien und die Fensterlänge gewählt.

6.4.3 Estes Modell

Die Simulationen des Estes Modells wurden mit einem konstanten Lernparameter und mit variablen Lernparametern für alle Wortpaare durchgeführt.

Zuerst wurde ein konstanter Lernparameter für alle Wortpaare gesucht, der das beste Ergebnis in Bezug auf die vier Kriterien lieferte. Zur Bestimmung des Lernparameters wurde der Hill-Climbing Algorithmus aus dem vorherigen Abschnitt verwendet. Bevor der Hill-Climbing Algorithmus verwendet werden konnte, mussten geeignete Startwerte für den Hill-Climbing Algorithmus ermittelt werden. Die unterschiedlichen Startwerte für den Hill-Climbing Algorithmus wurden durch eine schrittweise Verkleinerung des θ - Parameters ermittelt. Der Lernparameter θ im Estes Modell kann Werte zwischen 0 und 1 annehmen. Wurde der Wert für θ zu groß gewählt (z.B. 0.5), so wurde schon nach wenigen Lerndurchgängen die maximale Assoziationsstärke erreicht. Deswegen wurde als erster Wert für θ 0.05 gewählt. Dieser Wert wurde in den folgenden Simulationsreihen um 10er Potenzen verkleinert (s. Tabelle 6.3). Dann wurde ermittelt, bei welchem der obigen Werte für θ nach dem ersten Kriterium das beste Ergebnis erzielt wurde. In der Tabelle 6.3 werden die Ergebnisse, die mit den unterschiedlichen Werten für θ gewonnen wurden, aufgelistet.

θ	Kriterium 1	Fensterlänge
$5E - 2$	10	4
$5E - 3$	12	4,5,6,8
$5E - 4$	23	21,22
$5E - 5$	28	44,45
$5E - 6$	26	46-50
$5E - 7$	15	36-49
$1E - 7$	8	3

Tabelle 6.3: Die Ermittlung unterschiedlicher Startwerte für den Hill-Climbing Algorithmus. In der ersten Spalte steht der Wert, der für θ verwendet wurde. Die zweite Spalte enthält die Anzahl der richtig vorausgesagten Primärantworten, die mit dem jeweiligen Wert für θ erzielt wurden. In der dritten Spalte wird angegeben bei welchen Fensterlängen, die in der Spalte 2 angegebene Anzahl der richtig vorausgesagten Primärantworten erzielt wurden. Als Startwerte für den Hill-Climbing Algorithmus wurden die Werte $5E-5$ und $5E-6$ gewählt, da bei diesen Werten die meisten Primärantworten richtig vorausgesagt wurden.

Das beste Ergebnis in Bezug auf das erste Kriterium, 28 richtig vorausgesagte Primärantworten, wurde bei einem Wert von $5E - 5$ für θ erzielt. Dieses Ergebnis wurde bei einer Fensterlänge von 44 und 45 erreicht. Das zweitbeste Ergebnis lieferte der Wert $5E - 6$ für θ mit 26 richtig vorausgesagten Primärantworten bei den Fensterlängen von 46 bis 50. Diese beiden

Werte für den Parameter θ dienten nacheinander als Startwerte für einen iterativen Hill-Climbing Algorithmus. Als erstes wurde in der Umgebung von $5E - 5$ nach besseren Lösungen gesucht. Die Ergebnisse sind in der Tabelle 6.4 zusammengefasst.

θ	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$3.8E - 5$	28	254	37	52	29
$3.9E - 5$	28	254	36	53	28
$3.95E - 5$	28	254	37	53	28
$3.98E - 5$	28	254	37	53	28
$3.99E - 5$	28	254	37	53	28
$4E - 5$	28	254	37	53	28
$4.1E - 5$	27	263	36	55	15
$5.1E - 5$	27	171	30	38	33
$5.5E - 5$	27	171	30	37	34

Tabelle 6.4: Die erzielten Ergebnisse in Bezug auf die vier Kriterien in Abhängigkeit von θ , wenn in der Umgebung von $5E - 5$ mit dem Hill-Climbing Algorithmus nach optimaleren Lösungen gesucht wurde. Als Bewertungsfunktion dienten die vier Kriterien (Spalten 2-5), die nach ihrer Reihenfolge 1 - 4 gewichtet wurden. Ein θ -Parameter führte zu einer optimaleren Lösung, wenn bei gleicher Anzahl vorhergesagter Primärantworten dieses Ergebnis schon bei einer kleineren Fensterlänge erreicht wurde. Beste Ergebnisse in Bezug auf die vier Kriterien wurden um einen Wert von $4E - 5$ für θ herum gefunden.

Der iterative Hill-Climbing Algorithmus lieferte mit dem Startwert $5E - 5$ für θ bei einem Wert um $4E - 5$ für θ nur eine suboptimale Lösung. Denn wenn der Algorithmus mit dem Wert $5E - 6$ (eine 10er Potenz kleiner als der vorherige Startwert) gestartet wurde, wurden noch bessere Ergebnisse erzielt. Diese lagen bei einem Wert von $1E - 5$ für θ . Bei diesem Wert für θ wurden 30 Primärantworten bei einer Fensterlänge von 47 richtig vorausgesagt (siehe Tabelle 6.5).

Für das erste Kriterium wurde das beste Ergebnis bei einem θ von $1E - 5$ erzielt. Nimmt man Kriterium 2 hinzu, so ergaben sich für den Wert $2E - 5$ und $2.1E - 5$ die besseren Ergebnisse, wobei die Anzahl der richtig vorausgesagten Primärantworten um eins sank. Es wurden somit 29 Primärantworten bei einer Fensterlänge von 21 bzw. 20 richtig vorausgesagt, während beim zweiten Kriterium 274 bzw. 273 Nennungen der Teilnehmer gegenüber den 252 Nennungen der Teilnehmer mit dem Parameter $1E - 5$ richtig vorausgesagt werden konnten.

θ	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$1E - 5$	30	252	36	55	47
$1.01E - 5$	30	252	36	55	47
$1.05E - 5$	30	252	36	55	47
$1.1E - 5$	30	251	36	55	47
$1.7E - 5$	29	272	39	54	23
$1.8E - 5$	29	272	39	55	22
$1.9E - 5$	29	273	38	55	21
$2E - 5$	29	274	38	55	21
$2.1E - 5$	29	273	39	55	20
$2.2E - 5$	29	272	39	54	20
$2.3E - 5$	28	275	39	53	19

Tabelle 6.5: Die erzielten Ergebnisse in Bezug auf die vier Kriterien in Abhängigkeit von θ , wenn in der Nähe von $5E-6$ nach optimaleren Lösungen gesucht wurde. Das beste Ergebnis in Bezug auf das erste Kriterium wurde mit einem θ von $1E - 5$ und einer Fensterlänge von 47 erzielt. Dies waren 30 richtig vorausgesagte Primärantworten. Ab einem Wert von $1.7E - 5$ für θ wurde zwar eine Primärantwort weniger vorausgesagt, aber dafür verbesserten sich die Ergebnisse in Bezug auf das zweite und dritte Kriterium, und die Ergebnisse konnten schon bei einer Fensterlänge um 20 herum erzielt werden. Die besten Ergebnisse in Bezug auf alle vier Kriterien und minimaler Fensterlänge wurden bei einem θ von $1.7E - 5$ bis $2.2E - 5$ und bei den Fensterlänge von 20-23 erzielt.

Bei einer anschließenden visuellen Betrachtung der berechneten Wortassoziationen zu einigen Stimuluswörtern fiel direkt auf, dass die häufigsten Wörter aus dem Vokabular auch häufig auf den ersten zehn Rangplätzen zu finden waren, obwohl sie von keinem Teilnehmer genannt wurden. Um diesen Worthäufigkeitseffekt zu verringern, wurde der erste Lösungsansatz aus dem Abschnitt Worthäufigkeitseffekt verwendet. Die Assoziationsberechnungen, die mit dem Parameter $2.1E - 5$ ermittelt wurden, wurden anschließend mit Hilfe der Gleichung 6.2, die im Abschnitt Worthäufigkeitseffekt definiert wurde, gewichtet. Durch die Gewichtung wurden 30 Primärantworten richtig vorausgesagt, 39 Primärantworten wurden auf den Rängen 1 und 2, 58 Primärantworten wurden auf den ersten fünf Rängen richtig vorausgesagt, und 287 Nennungen (ohne die Primärantworten) aller Teilnehmer wurden auf den ersten zehn Rangplätzen gezählt. Dieses Ergebnis wurde bei einer Fensterlänge von 22 erzielt. Durch die Gewichtung konnten die Assoziationsberechnungen in Bezug auf die vier Kriterien leicht verbessert werden.

Der Worthäufigkeitseffekt wurde aber durch die Gewichtung nur sehr wenig verringert. Die häufigsten Wörter *sein*, *Seite* und *wollen* sanken in der Rangreihe zu einem Stimuluswort nur um ein Paar Rangplätze nach unten, so dass sie teilweise noch auf den ersten zehn Rangplätzen zu finden waren. Ein Beispiel hierfür sind die Assoziationen zu *Frau*, die in der Tabelle 6.6 aufgeführt werden.

Rang	Assoziationsstärke	Response	gewichtete Ass.	Response
1	0.064	Mann	0.673	Mann
2	0.035	jung	0.384	jung
3	0.030	sein	0.250	Kind
4	0.023	Kind	0.169	sein
5	0.012	wollen	0.149	Herr
6	0.012	Herr	0.1403	Mutter
7	0.011	gehen	0.1346	schön
8	0.011	schön	0.1265	katholisch
9	0.011	Mutter	0.1235	Sport
10	0.0108	gut	0.1148	Polizei

Tabelle 6.6: Die berechneten Assoziationen zu Frau mit dem Estes Modell. In der zweiten und dritten Spalte sind die nicht gewichteten Assoziationen aufgeführt, während in der vierten und fünften Spalte die Assoziationen mit dem Informationsgehalt des jeweiligen Responsewortes gewichtet wurden.

Das Wort *sein* fiel nur um einen Rangplatz, während das Wort *wollen* nicht mehr auf die ersten zehn Rangplätze kam.

Für die folgenden Simulationen soll ein variabler Lernparameter für alle Wortpaare verwendet werden. Bei diesen Simulationen soll der zweite Lösungsansatz zur Verringerung des Worthäufigkeitseffekts verwendet werden, der im Abschnitt Worthäufigkeitseffekt beschrieben wurde.

Der Lernparameter, der durch die Gleichung 6.3 im Abschnitt Worthäufigkeitseffekt spezifiziert wurde, enthält einen freien Parameter, der mit F bezeichnet wurde, der dafür sorgen sollte, dass der Lernparameter einen Wert zwischen 0 und 1 erhielt. Bei einem Faktor von $1E - 05$ erhielten alle Wortpaare einen Lernparameter, der im Wertebereich von 0 und 1 lag. Dieser Faktor $1E - 05$ wurde in den folgenden Simulationen schrittweise um $\frac{1}{10}$ verkleinert, um zu sehen, bei welchem Wert sich die besten Ergebnisse in Bezug auf die 4 Kriterien ergaben. Die Fensterlänge in einer Simulationsreihe wurde um zwei statt um eins erhöht, um Rechenzeit zu sparen. Eine Simulationsreihe wurde mit der Fensterlänge von 12 gestartet. Die folgenden Simulationen der Simulationsreihe wurden nur mit geraden Fensterlängen

durchgeführt, bis eine Fensterlänge von 100 erreicht wurde. Die Ergebnisse sind in der Tabelle 6.7 aufgeführt.

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$1E - 5$	13	104	14	21	20
$1E - 6$	19	211	32	46	20
$1E - 7$	28	284	38	57	22
$1E - 8$	23	215	36	47	20
$1E - 9$	18	183	21	40	18

Tabelle 6.7: Die erzielten Ergebnisse in Bezug auf die vier Kriterien bei Verwendung eines variablen Lernparameters für alle Wortpaare. Bei der Verwendung von $1E - 7$ als Faktor ergaben sich die besten Ergebnisse in Bezug auf die vier Kriterien.

Bei der visuellen Betrachtung der Assoziationen zu einigen Stimuluswörtern fiel auf, dass der Worthäufigkeitseffekt weiterhin vorhanden war. Dies lag daran, dass die unterschiedlichen Lernparameter für die einzelnen Wortpaare zu wenig untereinander variierten, da der Informationsgehalt eines Wortes nur im Wertebereich von 5 und 25 lag. Um eine größere Varianz bei den unterschiedlichen Lernparametern zu erhalten, wurde der Informationsgehalt der Wörter im Wortpaar potenziert.

$$\theta_{i,j} = F \cdot (I_i \cdot I_j)^n \quad (6.5)$$

Durch die Potenzierung des Informationsgehaltes erhielt man einen weiteren freien Parameter n , der ermittelt werden musste. Der zugehörige Skalierungsfaktor F hing natürlich von der gewählten Potenz n ab. Je größer die Potenz n gewählt wurde, umso kleiner musste der Wert für den F sein, damit der Lernparameter im Wertebereich zwischen 0 und 1 lag.

Es wurden Simulationen mit n gleich 2,3,4,5 und 6 durchgeführt. Eine Besserung in Bezug auf den Worthäufigkeitseffekt trat aber erst bei einer Potenzierung mit $n = 5$ auf. In Bezug auf das zweite Kriterium, bei dem alle Responsewörter der Teilnehmer auf den ersten zehn Rangplätzen gezählt wurden, ergab sich aber erst bei einer Potenzierung mit 5 eine sichtbare Verbesserung. Statt der 274 Responsewörter, die mit einem konstanten Lernparameter von 0.000021 erzielt wurden, wurden über 300 Responsewörter auf den ersten zehn Rangplätzen gezählt. Bei den anderen Kriterien ergaben sich keine Verbesserungen.

Das beste Ergebnis in Bezug auf die Kriterien 1 und 2 und minimaler Fensterlänge wurde bei einer Potenzierung mit $n = 5$ und einem Skalierungsfaktor von $3.6E - 19$ erzielt. Bei einer Fensterlänge von 22 wurden 29 richtig

vorausgesagte Primärantworten auf dem Rangplatz 1 gezählt, 309 richtig vorausgesagte Responsewörter der Teilnehmer auf den ersten 10 Rangplätzen, 38 richtig vorausgesagte Primärantworten auf den ersten beiden Rangplätzen und 55 richtig vorausgesagte Primärantworten auf den ersten fünf Rangplätzen. Bei der Potenzierung mit $n = 5$ konnten auch 30 Primärantworten richtig vorausgesagt werden. Dies war aber erst bei einer Fensterlänge von 52 möglich. Das Kriterium 2 verschlechterte sich dann aber auf 296 richtig vorausgesagte Responsewörter der Teilnehmer auf den ersten zehn Rangplätzen.

Bei einer Potenzierung mit $n = 6$ konnten sogar 323 Responsewörter der Teilnehmer auf den ersten 10 Rangplätzen richtig vorausgesagt werden, die Anzahl der Primärantworten, die richtig vorausgesagt wurden, sank aber dann auf 27. Auch konnten die 323 Responsewörter der Teilnehmer erst mit einer Fensterlänge von 42 vorausgesagt werden. Mit der Potenzierung von 6 war sogar eine Voraussage von 31 Primärantworten möglich. Somit lieferte eine Potenzierung mit $n=6$ das beste Ergebnis in Bezug auf das Kriterium 1. Dies war aber wieder, wie bei der Potenzierung mit 5, erst bei einer größeren Fensterlänge (ab 68) möglich. Wie bei der Potenzierung mit 5 konnten dann beim Kriterium 2 weniger Responsewörter (291) auf den ersten zehn Rangplätzen vorausgesagt werden.

Aufgrund der vielen Möglichkeiten für die Wahl des Skalierungsfaktors konnten nicht alle möglichen Werte ausprobiert werden, sondern es wurden für jede untersuchte Potenzierung nur Stichproben durchgeführt. Bei der Potenzierung mit 5 wurden die meisten unterschiedlichen Faktoren ausprobiert, da der Worthäufigkeitseffekt hier sichtbar verringert werden konnte und da in Bezug auf das zweite Kriterium größere Verbesserungen eintraten.

Im folgenden Abschnitt werden die durchgeführten Simulationen in Tabellenform aufgeführt. Für jede Potenzierung von n gleich 2,4,5 und 6 gibt es eine Tabelle. Die erste Spalte enthält den gewählten Faktor F , die Spalten 2 bis 5 enthalten die Ergebnisse in Bezug auf die vier Kriterien und die letzte Spalte enthält die Fensterlänge, bei der dieses Ergebnis erzielt wurde. Alle Simulationsreihen wurden mit einer Fensterlänge von 20 gestartet und für die folgenden Simulationen wurde diese immer um 2 erhöht. Wenn die Ergebnisse bei einer Fensterlänge in Bezug auf das Kriterium 1 und das Kriterium 2 schlecht waren, so wurden keine weiteren Simulationen mit diesem Faktor durchgeführt, d.h. die Simulationsreihe wurde abgebrochen. Schlecht in Bezug auf das erste Kriterium bedeutete: Es wurden nicht mindestens 20 Primärantworten auf Rangplatz 1 vorausgesagt. Die Ergebnisse, die in Bezug auf das Kriterium 2 als schlecht bewertet wurden, waren die, bei denen nicht mindestens 250 Nennungen der Teilnehmer auf den ersten zehn Rangplätzen vorausgesagt werden konnten. Wurden ab einer bestimmten Fensterlänge die Ergebnisse in Bezug auf erste oder das zweite Kriterium stetig schlechter, so

wurden ebenfalls keine weiteren Simulationen mehr in dieser Simulationsreihe durchgeführt.

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$1E - 8$	19	164	38	39	20
$1E - 9$	27	265	38	53	22
$1E - 10$	26	252	39	55	26
$2E - 10$	27	272	35	58	26
$3.5E - 10$	28	284	36	57	22
$5E - 10$	29	255	39	49	60
$6E - 10$	29	267	38	51	38
$2E - 11$	18	221	31	47	22

Tabelle 6.8: Estes Modell: Ergebnisse, die mit dem Parameter $\theta_{i,j} = F \cdot (I_i \cdot I_j)^2$ erzielt wurden.

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$1E - 11$	23	240	37	51	20
$1E - 12$	28	280	36	55	34
$1.5E - 12$	29	284	36	55	30
$2.1E - 12$	28	279	38	57	24
$1E - 13$	24	253	35	53	24

Tabelle 6.9: Estes Modell: Ergebnisse, die mit dem Parameter $\theta_{i,j} = F \cdot (I_i \cdot I_j)^3$ erzielt wurden.

In Bezug auf den Worthäufigkeitseffekt wurden keine Verbesserungen bei Potenzierungen mit $n=2$ (siehe Tabelle 6.8), $n=3$ (siehe Tabelle 6.9) und $n=4$ (siehe Tabelle 6.10) erzielt. Erst ab einer Potenzierung mit $n=5$ (siehe Tabelle 6.11) fiel der Worthäufigkeitseffekt nicht mehr so stark ins Gewicht.

Ab einem Faktor von $3.6E - 19$ war das häufigste Wort *sein* bei den Stimuluswörtern *Mann* und *Frau* nicht mehr auf den ersten zehn Rangplätzen zu finden. Mit diesem Faktor wurden die besten Ergebnisse in Bezug auf die vier Kriterien erzielt (siehe Tabelle 6.11).

Bei einer Potenzierung mit 6 und einem Faktor von $5E - 20$ wurden die meisten Primärantworten richtig vorausgesagt. Dies waren 31 richtig vorausgesagte Primärantworten (siehe Tabelle 6.12). Der Worthäufigkeitseffekt konnte mit diesem Faktor und dieser Potenzierung deutlich verringert werden. Dies kann man z.B. an den berechneten Assoziationen zu dem Stimuluswort *Frau* sehen, die in der Tabelle 6.13 aufgeführt sind, sehen.

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
1E – 13	20	207	31	46	20
1E – 14	27	276	37	55	24
1E – 15	28	285	36	57	22
1E – 16	28	274	37	54	34
1E – 17	29	279	38	55	34
1E – 18	29	280	38	57	34
5E – 18	29	276	38	55	34
1E – 19	28	292	36	51	32
1E – 20	24	298	35	49	26

Tabelle 6.10: Estes Modell: Die erzielten Ergebnisse in Bezug auf die vier Kriterien, die mit dem Parameter $\theta_{i,j} = F \cdot (I_i \cdot I_j)^4$ erzielt wurden. Bei einem Faktor von 1E – 19 verbesserten sich die Ergebnisse in Bezug auf das zweite Kriterium leicht.

Bei einer Potenzierung mit 5 oder mit 6 ergaben sich die besten Ergebnisse in Bezug auf die vier Gütekriterien. Für die bisherigen Simulationen diente das Textkorpus SZ98 als Kalibrierkorpus zur Bestimmung des freien Parameters θ . Für die Simulationen mit den Testkorpora wurden folgende Parameter ausgewählt:

Ein konstanter Parameter für alle Wortpaare. Dies war der Parameter $\theta = 0.000021$, da mit diesem Parameter die besten Ergebnisse in Bezug auf die vier Gütekriterien erzielt wurden, wenn nur konstante Parameter verwendet wurden. Eine anschließende Gewichtung der berechneten Ergebnisse mit dem Informationsgehalt konnte die Ergebnisse, die mit $\theta = 0.000021$ gewonnen wurden, verbessern. Diese Gewichtung sollte auch bei den Simulationen mit den Testkorpora durchgeführt werden. Bei einem nicht konstantem θ wurden die besten Ergebnisse mit folgenden beiden Parametern erzielt:

$$\theta_{i,j} = 3.6E - 19 \cdot (I_i \cdot I_j)^5 \quad (6.6)$$

$$\theta_{i,j} = 5E - 20 \cdot (I_i \cdot I_j)^6 \quad (6.7)$$

Der Parameter, der durch die Gleichung 6.6 beschrieben wird, lieferte gute Ergebnisse in Bezug auf die ersten beiden Kriterien schon ab einer Fensterlänge von 22. Eine Primärantwort mehr wurde mit gleichem Parameter bei einer Fensterlänge von über 50 erzielt. Mit dem Parameter, der durch die Gleichung 6.7 spezifiziert wird, konnten sogar 31 Primärantworten richtig vorausgesagt werden. Dies war aber erst bei einer Fensterlänge von weit

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$3.6E - 17$	27	286	38	55	24
$1.6E - 18$	29	307	38	55	22
$1.6E - 18$	30	299	37	53	60
$3E - 18$	29	307	38	55	22
$3.6E - 18$	29	308	38	57	22
$3.6E - 18$	30	296	37	53	58
$4E - 18$	29	307	38	57	22
$1E - 19$	29	306	38	55	22
$1E - 19$	30	296	37	53	58
$2.6E - 19$	29	308	38	55	22
$3.6E - 19$	29	309	38	55	22
$3.6E - 19$	30	302	38	54	62
$1.2E - 20$	29	305	38	58	22
$2E - 22$	27	298	37	50	22

Tabelle 6.11: Estes Modell: Ergebnisse in Bezug auf die vier Kriterien, die mit dem Parameter $\theta_{i,j} = F \cdot (I_i \cdot I_j)^5$ erzielt wurden. Bei einem Faktor aus dem Bereich von $1.6E - 18$ bis $1.2E - 20$ ergaben sich fast gleich gute Ergebnisse in Bezug auf alle Kriterien. Bei den Faktoren $1.6E - 18$, $3.6E - 18$, $1E - 19$ und $3.6E - 19$ sind zwei Ergebnisse aufgeführt, da die maximale Anzahl richtiger Primärantworten erst bei einer Fensterlänge von 58 bzw. 62 erreicht wurde. Die Ergebnisse in Bezug auf das Kriterium 2 fielen aber bei den Fensterlängen von 58 bzw. 62 etwas schlechter aus als bei einer Fensterlänge von 22. Bei einer Fensterlänge von 22 wurde aber nur eine Primärantwort weniger vorausgesagt als bei einer Fensterlänge von 58 bzw. 62.

über 50 möglich. Das zweite Kriterium verschlechterte sich aber bei steigender Fensterlänge. Es fiel dann von 309 Nennungen bei einer Fensterlänge von 38 auf 291 Nennungen bei einer Fensterlänge von 68. Deswegen wurde der Parameter, der durch die Gleichung 6.6 spezifiziert wird, als der bessere von den oberen beiden Parametern bewertet.

Für die Simulationen mit den Testkorpora wurde somit der Parameter, der durch die Gleichung 6.6 beschrieben wird, als nicht konstanter Parameter für alle Wortpaare ausgewählt. Alle ausgewählten Parameter lieferten schon mit einer Fensterlänge von 22 gute Ergebnisse in Bezug auf die vier Kriterien.

In den beiden Abbildungen 6.1 und 6.2 werden die Ergebnisse der drei ausgewählten Parameter in Abhängigkeit von der Fensterlänge dargestellt. Die Fensterlänge variiert in beiden Abbildungen von 12 bis 50. Es werden die Ergebnisse in Bezug auf das erste und das zweite Kriterium dargestellt,

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$1.6E - 19$	27	290	38	53	26
$1.6E - 20$	27	320	39	53	44
$1.6E - 20$	29	296	39	52	82
$5E - 20$	28	309	39	52	38
$5E - 20$	31	291	39	51	68
$8E - 20$	28	295	40	53	40
$8E - 20$	30	291	38	52	58
$3.6E - 21$	27	320	39	53	46
$3.6E - 21$	29	302	40	54	92
$2E - 22$	27	322	39	53	44
$1E - 23$	27	314	39	54	26
$1E - 24$	25	302	36	47	32

Tabelle 6.12: Estes Modell: Ergebnisse in Bezug auf die vier Kriterien, die mit dem Parameter $\theta_{i,j} = F \cdot (I_i \cdot I_j)^6$ erzielt wurden. Bei den Faktoren $1.6E - 20$, $5E - 20$, $8E - 20$ und $3.6E - 21$ sind zwei Ergebnisse angegeben. Das erste Ergebnis gibt das beste Ergebnis in Bezug auf das Kriterium 1 und Kriterium 2 und minimaler Fensterlänge an. Während das zweite angegebene Ergebnis das beste Ergebnis in Bezug auf das Kriterium 1 ist.

da die beiden ersten Kriterien den größten Einfluss auf die Parameterwahl hatten. In der Abbildung 6.1 werden die Ergebnisse in Bezug auf das erste Kriterium graphisch dargestellt. Die Kurve, die durch Dreiecke gekennzeichnet wird, gehört zu den Ergebnissen, die mit einem variablen Parameter für alle Wortpaare gewonnen wurden. Die Kurve mit den Rechtecken steht für die Ergebnisse, die durch Verwendung eines konstanten Parameters erzielt wurden. Wurden die Ergebnisse, die zur Kurve mit den Rechtecken gehören, anschließend noch mit dem Informationsgehalt des jeweiligen Responsewortes gewichtet, so entstanden die Ergebnisse, die die Kurve mit Rauten darstellt. Die Anzahl der richtig vorhergesagten Primärantworten, die mit einem variablen Parameter (Kurve mit Dreiecken) erzielt wurden, blieb ab einer Fensterlänge von 18 fast konstant (28 oder 29 Primärantworten). Bei Verwendung eines konstanten Parameters (Kurve mit Rechtecken) wurde das Maximum (29 Primärantworten) bei Fensterlängen um die 22 erreicht. Danach fiel die Anzahl richtiger Primärantworten leicht ab. Bei einer Fensterlänge von 46 war nochmal ein leichter Anstieg bei der Anzahl der richtig vorausgesagten Primärantworten zu beobachten, bevor die Anzahl der richtig vorausgesagten Primärantworten bei einer Fensterlänge von 50 auf 25 fiel. Wurden die Ergebnisse, die mit einem konstanten Parameter erzielt worden waren, anschließend

noch gewichtet (Kurve mit Rauten), so wurde bei einer Fensterlänge von 22 das Maximum mit 30 richtig vorausgesagten Primärantworten erzielt. Dies war eine Primärantwort mehr als bei den anderen beiden Parametern. Durch die Gewichtung fiel die Anzahl der richtig vorausgesagten Primärantworten mit einer Fensterlänge von 50 nicht auf 25, sondern sie stieg wieder auf 29 an, nachdem sie zuvor bei den Fensterlängen von 42, 44 und 46 auf 27 abgefallen war.

In der Abbildung 6.2 werden die Ergebnisse in Bezug auf das zweite Kriterium in Abhängigkeit von der Fensterlänge dargestellt. Bei Verwendung eines variablen Parameters für alle Wortpaare (Kurve mit Dreiecken) wurden bei allen dargestellten Fensterlängen die besten Ergebnisse (von den drei ausgewählten Parametern) erzielt. Am schlechtesten waren die Ergebnisse, wenn ein konstanter Parameter für alle Wortpaare (Kurve mit Rechtecken) verwendet wurde. Etwas bessere Ergebnisse konnten durch eine anschließende Gewichtung (Kurve mit Rauten) erzielt werden.

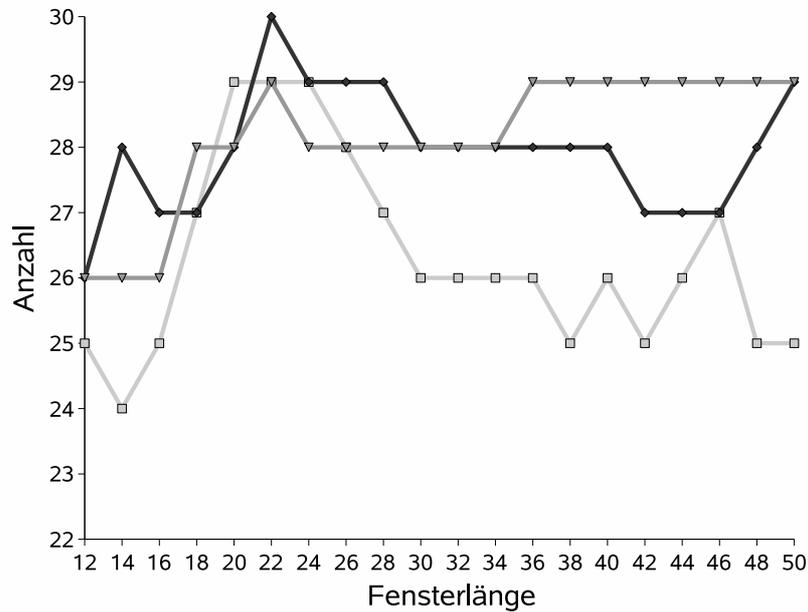


Abbildung 6.1: Die Anzahl der richtig vorhergesagten Primärantworten (Kriterium 1), die mit den drei Varianten des Estes Modells in Abhängigkeit von der Fensterlänge erzielt wurden.

Die Kurve mit den Dreiecken repräsentiert die richtig vorausgesagten Primärantworten mit dem Estes Modell, wenn ein variabler Lernparameter für die einzelnen Wortpaare verwendet wurde. Die Kurve mit den Rechtecken entstand, wenn ein konstanter Lernparameter für alle Wortpaare bei den Assoziationsberechnungen verwendet wurde. Die Kurve mit den Rauten gehört zu den Ergebnissen, die durch eine anschließende Gewichtung mit dem Informationsgehalt des jeweiligen Responsewortes entstanden worden waren. Die Gewichtung wurde mit den Ergebnissen, die mit einem konstanten Lernparameter entstanden worden waren, durchgeführt.

Alle Parameter erzielten die meisten Primärantworten mit einer Fensterlänge von 22.

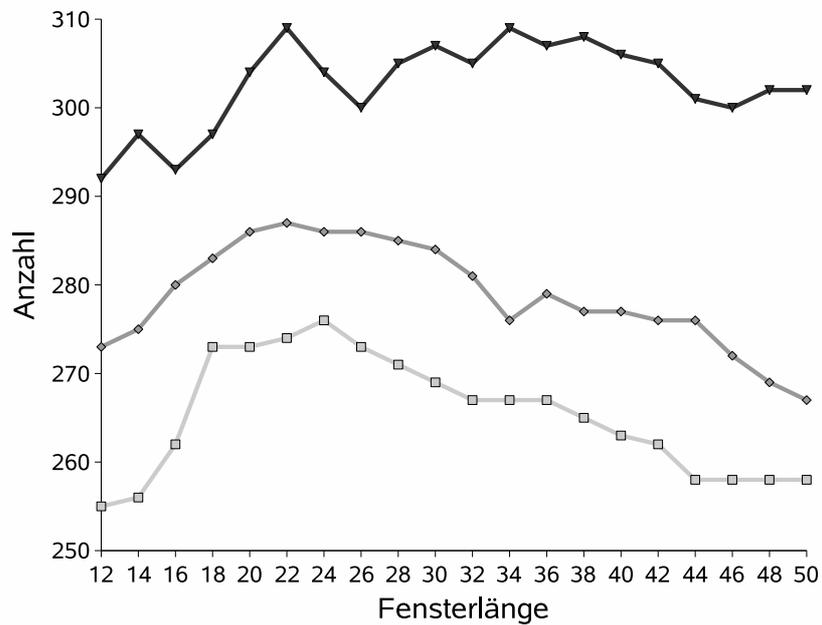


Abbildung 6.2: Die Ergebnisse der drei Varianten des Estes Modells in Bezug auf das Kriterium 2 in Abhängigkeit von der Fensterlänge.

Die obere Kurve mit den Dreiecken gehört zu den Ergebnissen, die mit einem variablen Parameter erzielt wurden. Die untere Kurve mit den Rechtecken repräsentiert die Ergebnisse, die mit einem konstanten Parameter für alle Wortpaare erzielt wurden. Die mittlere Kurve mit den Rauten entstand durch eine anschließende Gewichtung mit dem Informationsgehalt des jeweiligen Responsewortes. Die Gewichtung erfolgte mit den Ergebnissen, die durch die untere Kurve repräsentiert sind.

Alle Parameter lieferten mit einer Fensterlänge von 22 gute Ergebnisse in Bezug auf das zweite Kriterium.

Rangplatz	Assoziationsstärke	Responsewort
1	0.04150	Mann
2	0.01476	Kind
3	0.01449	Mutter
4	0.01410	jung
5	0.01073	Mädchen
6	0.01005	Polizei
7	0.00956	Schweiz
8	0.00846	Liebe
9	0.00841	schön
10	0.00799	Haus

Tabelle 6.13: Estes Modell: Assoziationen zu dem Stimuluswort *Frau*, die mit dem Parameter $\theta_{i,j} = 5E - 20 \cdot (I_i \cdot I_j)^6$ berechnet wurden. Der Worthäufigkeitseffekt konnte bei den Assoziationen zu *Frau* deutlich verringert werden. Die sehr häufigen Wörter *sein*, *wollen*, *gut* und *gehen* wurden nicht mehr auf den ersten zehn Rangplätzen als Responsewörter zu dem Stimuluswort *Frau* ausgegeben.

Das Wort *Polizei* wurde so stark mit *Frau* und *Mann* assoziiert, da in den verwendeten Zeitungstexten häufig Berichte von Verbrechen und Festnahmen durch die Polizei zu lesen waren, in denen Männer und Frauen zusammen mit der Polizei vorkamen. Die starke Assoziationsstärke von *Frau* und *Schweiz* entstand durch die häufigen Berichte von Sportwettkämpfen in den Zeitungen.

In den Ergebnistabellen von Sportwettkämpfen kommt erst der Ländername und dann meistens direkt danach Männer oder Frauen. Da es sich bei den Sportwettkämpfen in der Schweiz häufig um Wintersport (Ski) handelt, sind die Ergebnistabellen, die Schweiz und Frau enthalten, häufig am Ende der Textkorpora zu finden, da am Ende des Jahres viele Wettkämpfe im Wintersport ausgetragen werden. Da die Assoziation von Schweiz und Frau am Ende aufgebaut wurde, konnte sie nicht durch einen folgenden Text wieder abgebaut werden.

6.4.4 Rescorla & Wagner Modell

Das Rescorla & Wagner Modell hat drei freie Parameter. Dadurch ergaben sich sehr viele Kombinationsmöglichkeiten für die Wahl der freien Parameter. Da die Rechenzeit für das Rescorla & Wagner Modell gegenüber dem Estes Modell um einen Faktor von ca. 10 größer war, konnten nicht so viele Simulationen, wie mit dem Estes Modell, mit dem Rescorla & Wagner Modell durchgeführt werden. Bei einer Fensterlänge von 20 benötigte das Estes Modell ca. 20 Minuten Rechenzeit, während das Rescorla & Wagner Modell 3 Stunden und 20 Minuten benötigte. Die Rechenzeit für das Estes Modell variierte bei zunehmender Fensterlänge nicht so stark wie beim Rescorla & Wagner Modell. Sie lag beim Estes Modell zwischen 20 bis 30 Minuten (30 Minuten bei einer Fensterlänge von 100), während beim Rescorla & Wagner Modell bei einer Fensterlänge von 32 schon 5 Stunden benötigt wurden.

Die höhere Rechenzeit beim Rescorla & Wagner Modell entstand dadurch, dass beim Vergessen jedes mal die zugehörigen Einträge in der Matrix neu berechnet werden mussten, während beim Estes Modell mit der neuen Berechnung bis zum nächsten Lernen gewartet werden konnte. Es musste nur gezählt werden, wie häufig die Regel für Vergessen auf das jeweilige Wortpaar anzuwenden war. Diese Möglichkeit der Rechenzeitverbesserung konnte beim Rescorla & Wagner Modell nicht verwendet werden, da die Änderung der Assoziationsstärke in jedem Fenster variierte. Sie ist abhängig von den anderen Wörtern im Fenster und musste deshalb nach jeder Fensterverschiebung neu bestimmt werden.

Da die Assoziationsmatrix beim Rescorla & Wagner Modell nicht symmetrisch ist, wurde für das Rescorla & Wagner Modell doppelt soviel Speicherplatz wie für das Estes Modell benötigt.

Um die Kombinationsmöglichkeiten einzuschränken, wurde für λ , die maximale Assoziationsstärke, ein Wert von 1 gewählt. Durch diese Einschränkung reduzierte sich die Anzahl der freien Parameter im Rescorla & Wagner Modell von drei auf zwei.

Die Simulationsreihen wurden, wie beim Estes Modell, mit einer Fensterlänge von 12 gestartet, und danach wurde in den folgenden Simulationen die Fensterlänge iterativ um 2 erhöht. Da die Rechenzeit für das Rescorla & Wagner Modell so hoch war, konnten die Simulationsreihen nur bis zu einer maximalen Fensterlänge von 50 durchgeführt werden. Eine Simulationsreihe wurde nicht bis zu einer Fensterlänge von 50 durchgeführt, wenn sich die Ergebnisse in Bezug auf eines der ersten beiden Kriterien, Kriterium 1 oder Kriterium 2, stark verschlechterten.

Für die ersten Simulationsreihen wurden α und β zu einem Parameter zusammengefasst, dessen Wertebereich zwischen 0 und 1 lag. Zuerst wur-

de dieser Parameter, wie beim Estes Modell, für alle Wortpaare konstant gehalten. Für das Lernen und das Vergessen wurden dieselben Parameter verwendet. Die Simulationsreihen wurden mit einem Parameter von 0.02 gestartet, der in den folgenden Simulationsreihen schrittweise um $\frac{1}{10}$ verkleinert wurde (von 0.02 bis 0.0000002).

Diese Parameter lieferten keine geeigneten Startpunkte für den Hill-Climbing Algorithmus. Bei allen 6 Parametern wurden nur sehr schlechte Ergebnisse in Bezug auf alle vier Kriterien erzielt. Bei den berechneten Assoziationen war ein sehr starker Worthäufigkeitseffekt zu beobachten. Durch diesen starken Worthäufigkeitseffekt konnten nur maximal 6 Primärantworten richtig vorausgesagt werden. Beim Estes Modell war der Worthäufigkeitseffekt nicht so stark ausgeprägt, wenn eine konstanter Parameter verwendet wurde. Da die mit einem konstanten Parameter erzielten Ergebnisse beim Rescorla & Wagner Modell so schlecht in Bezug auf die vier Kriterien ausgefallen waren, wurde der erste Lösungsansatz zum Verringern des Worthäufigkeitseffekts, der im Abschnitt Worthäufigkeitseffekt beschrieben wurde, nicht durchgeführt. Denn der erste Lösungsansatz konnte beim Estes Modell den Worthäufigkeitseffekt nur wenig verringern.

Für die nächsten Simulationsreihen wurde der zweite Lösungsansatz zur Verringerung des Worthäufigkeitseffekts verwendet. Die Parameter α_i und β_j entsprachen dann dem $\theta_{i,j}$ aus dem Estes Modell für jedes Wortpaar. Es war zu erwarten, dass der Informationsgehalt alleine nicht ausreichen würde, um den Worthäufigkeitseffekt zu verringern. Deswegen wurde, wie beim Estes Modell, der Informationsgehalt der Wörter im Wortpaar noch potenziert.

$$\alpha_i \cdot \beta_j = F \cdot (I_i \cdot I_j)^n \quad (6.8)$$

Der Faktor F in der Gleichung 6.8 musste so gewählt werden, dass der Parameter $\alpha_i \cdot \beta_j$ im Wertebereich von 0 und 1 lag.

Beim Estes Modell wurden mit einer Potenzierung von 5 ($n=5$) und von 6 ($n=6$) die besten Ergebnisse erzielt. Deshalb wurden die folgenden Simulationsreihen des Rescorla & Wagner Modells auch mit diesen Potenzierungen gestartet. In der Tabelle 6.14 sind die Ergebnisse in Bezug auf die vier Kriterien bei einer Potenzierung von 5 aufgelistet. Bei dieser Potenzierung wurden 27 richtig vorausgesagte Primärantworten erzielt. Eine Potenzierung mit 6 lieferte eine Primärantwort mehr und leichte Verbesserungen beim dritten Kriterium.

Die Ergebnisse, die sich bei den Simulationsreihen mit einer Potenzierung von 6 ergaben, sind in der Tabelle 6.15 aufgeführt.

Beim Rescorla & Wagner Modell konnten die Parameter für Lernen und Vergessen unterschiedlich gewählt werden. Nach Cheng (Cheng, 1997) sollten

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$3.6E - 16$	20	216	32	45	12
$3.6E - 17$	27	271	34	54	16
$1E - 18$	26	285	36	51	22
$3.6E - 18$	26	288	37	51	22
$3.6E - 19$	26	287	36	51	22
$3.6E - 21$	26	287	36	51	22
$3.6E - 22$	26	287	36	51	22
$3.6E - 23$	26	287	36	51	22

Tabelle 6.14: Rescorla & Wagner Modell: Ergebnisse in Bezug auf die vier Kriterien, die mit dem Parameter $\alpha_i \cdot \beta_j = F \cdot (I_i \cdot I_j)^5$ erzielt wurden, in Abhängigkeit von den unterschiedlichen Faktoren. Das beste Ergebnis in Bezug auf das erste Kriterium, die Anzahl richtig vorausgesagter Primärantworten, wurde mit einem Faktor von $F = 3.6E - 17$ erzielt. Mit diesem Faktor wurden 27 Primärantworten richtig vorausgesagt. Wurden kleinere Faktoren verwendet, kleiner oder gleich dem Faktor $F = 1E - 18$, so verbesserten sich die Ergebnisse in Bezug auf das zweite Kriterium nur leicht, während die Anzahl richtig vorausgesagter Primärantworten um eins sank.

sie dies, da für β beim Auftreten des US normalerweise ein größerer Wert angenommen wird, als wenn der US nicht auftritt. Die unterschiedlichen Parameter für Lernen und Vergessen wurden bei den folgenden Simulationsreihen dadurch realisiert, dass der Faktor F in Gleichung 6.8 beim Vergessen kleiner gewählt wurde als beim Lernen. Für die Simulationsreihen mit unterschiedlichen Faktoren wurden wieder die beiden Potenzierungen 5 und 6 verwendet. Die Tabelle 6.16 enthält die Ergebnisse, die sich mit einer Potenzierung von 5 ergaben. Durch die Wahl von unterschiedlichen Faktoren F bei einer Potenzierung von 5 wurden keine besseren Ergebnisse in Bezug auf die vier Kriterien erzielt, als wenn gleiche Faktoren bei derselben Potenzierung verwendet wurden.

Bei einer Potenzierung von 6 (siehe Tabelle 6.17) wurden ebenfalls keine besseren Ergebnisse in Bezug auf die vier Kriterien erzielt, als bei der Wahl von gleichen Faktoren bei derselben Potenzierung.

Da die Wahl von unterschiedlichen Faktoren zu keiner Verbesserung der Ergebnisse in Bezug auf die vier Kriterien geführt hatte, wurden in den nächsten Simulationsreihen wieder derselbe Faktor für Lernen und Vergessen verwendet.

Unterschiedliche Parameter für Lernen und Vergessen erhält man auch

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$3.6E - 18$	21	184	28	42	12
$3.6E - 19$	26	259	34	49	22
$1E - 20$	27	288	39	50	22
$3.6E - 20$	28	283	39	50	22
$5.6E - 20$	28	283	38	51	22
$9E - 20$	28	279	37	51	22
$3.6E - 21$	27	289	39	50	22
$3.6E - 22$	26	288	40	51	24

Tabelle 6.15: Rescorla & Wagner Modell: Ergebnisse in Bezug auf die vier Kriterien, die mit dem Parameter $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^6$ erzielt wurden, in Abhängigkeit von den unterschiedlichen Faktoren.

durch eine unterschiedliche Potenzierung bei den Parametern für Lernen und Vergessen. Beim Lernen wurde die Potenzierung größer gewählt als beim Vergessen. Durch eine Multiplikation mit gleichem Faktor entstand ein Lernparameter der größer war als der Parameter für das Vergessen. Die erzielten Ergebnisse mit diesen Parametern werden in der Tabelle 6.18 dargestellt.

Die verschiedene Parameter für Lernen und Vergessen hatten nicht zu besseren Ergebnissen in Bezug auf die vier Kriterien geführt.

In den folgendem Abschnitt werden Simulationsreihen beschrieben, die nur noch die Lernregel verwenden, d.h. der Parameter β im Rescorla & Wagner Modell wurde für jedes Wortpaar auf Null gesetzt. Das Weglassen der Regel für Vergessen führte zu einer enormen Beschleunigung der Rechenzeit. Bei einer Fensterlänge von 20 war eine Simulation schon nach 4 Minuten beendet, während mit der Regel für Vergessen 3 Stunden und 20 Minuten benötigt wurden. Auch die Simulationsreihen ohne Regel für Vergessen wurden wieder mit den unterschiedlichen Potenzierungen durchgeführt.

In den Tabellen 6.19 und 6.20 werden die Ergebnisse in Bezug auf die vier Kriterien dargestellt, wenn nur die Lernregel aus dem Rescorla & Wagner Modell verwendet wurde. Die Ergebnisse, die in den beiden Tabellen dargestellt werden, unterscheiden sich in Bezug auf die vier Kriterien kaum voneinander. Dies bedeutet, dass die Wahl des Faktors einen sehr geringen Einfluss auf die Ergebnisse hatte. Die Ergebnisse in Bezug auf das zweite Kriterium konnten durch die Potenzierung mit 5 gegenüber einer Potenzierung mit 4 verbessert werden. Bei dem ersten Kriterium wurde bei einer Potenzierung mit 5 nur eine Primärantwort mehr vorausgesagt als bei einer Potenzierung mit 4.

Eine Potenzierung mit 6 (siehe Tabelle 6.21) führte zu keiner Verbesserung in Bezug auf die vier Kriterien.

L	V	K 1	K 2	K 3	K 4	Fensterlänge
$3.6E - 15$	$3.6E - 17$	27	272	34	54	16
$3.6E - 15$	$3.6E - 18$	26	288	36	51	22
$6E - 15$	$6E - 18$	26	284	37	52	22
$3.6E - 16$	$3.6E - 17$	27	271	34	54	16
$3.6E - 17$	$3.6E - 18$	26	287	37	51	22

Tabelle 6.16:

Rescorla & Wagner Modell: Ergebnisse, die mit den unterschiedlichen Parametern für Lernen und Vergessen erzielt wurden.

$$\text{Lernparameter: } \alpha_i \beta_j = L \cdot (I_i \cdot I_j)^5$$

$$\text{Parameter für Vergessen: } \alpha_i \beta_j = V \cdot (I_i \cdot (I_j)^5)$$

Die unterschiedlichen Parameter für Lernen und Vergessen entstanden durch die Wahl von unterschiedlichen Faktoren. In der ersten Spalte ist der Faktor für Lernen (L) aufgeführt, während die zweite Spalten den Faktor für Vergessen (V) enthält.

In der Tabelle 6.22 werden die Ergebnisse präsentiert, die bei einer Potenzierung von 5.5 erzielt wurden.

Bei der Potenzierung mit 5.5 erhöhte sich die Anzahl richtig vorausgesagter Primärantworten um 2 gegenüber den beiden Potenzierungen mit 5 und mit 6. Bei allen untersuchten Potenzierungen hatte der Faktor F , der in der ersten Spalte aufgeführt wird, einen sehr geringen Einfluss auf die Ergebnisse in Bezug auf die vier Kriterien.

Bei den bisherigen Simulationen des Rescorla & Wagner Modells konnten ohne die Anwendung der Regel für Vergessen gleich gute Ergebnisse in Bezug auf die vier Kriterien erzielt werden wie mit der Anwendung der Regel für Vergessen. Die bisher erzielten Ergebnisse in Bezug auf die vier Kriterien waren aber alle schlechter als die Ergebnisse, die mit dem Estes Modell erzielt wurden. Insbesondere beim zweiten Kriterium wurde beim Estes Modell ein Wert über 300 erreicht, der bisher beim Rescorla & Wagner Modell noch nicht erzielt werden konnte.

Eine mögliche Ursache für das schlechtere Abschneiden des Rescorla & Wagner Modells beim zweiten Kriterium könnte die unterschiedliche Art der Wortpaarbildung sein. Im dritten Kapitel wurden zwei Varianten der Wortpaarbildung beschrieben (siehe Abschnitt Fenstertechnik). Für die Simulation des Estes Modells wurde die 1. Variante verwendet, während für das Rescorla & Wagner Modell die 2. Variante verwendet wurde. Durch die unterschiedlichen Varianten der Wortpaarbildung wurde beim Estes Modell die Assoziationsstärke von jedem Wort zu jedem anderem Wort im Fenster

Lernen L	Vergessen V	K 1	K 2	K 3	K 4	Fensterlänge
$3.6E - 18$	$3.6E - 19$	26	258	34	49	20
$3.6E - 18$	$3.6E - 20$	28	283	39	50	22
$6E - 18$	$6E - 20$	28	283	39	51	22
$9E - 18$	$9E - 20$	28	278	37	51	22
$3.6E - 18$	$3.6E - 21$	27	289	39	50	22
$3.6E - 19$	$3.6E - 21$	27	289	39	50	22
$1E - 20$	$1E - 21$	26	287	40	51	24

Tabelle 6.17:

Rescorla & Wagner Modell: Ergebnisse, die mit den unterschiedlichen Parametern für Lernen und Vergessen erzielt wurden.

$$\text{Lernparameter: } \alpha_i \beta_j = L \cdot (I_i \cdot I_j)^6$$

$$\text{Parameter für Vergessen: } \alpha_i \beta_j = V \cdot (I_i \cdot I_j)^6$$

Die unterschiedlichen Parameter für Lernen und Vergessen wurden durch die Wahl von unterschiedlichen Faktoren realisiert. In der ersten Spalte ist der Faktor für Lernen aufgeführt, während die zweite Spalte den Faktor für Vergessen enthält.

erhöht, während beim Rescorla & Wagner Modell nur die Assoziationsstärken der Wörter, die ein CS repräsentierten, zu einem Wort, welches den US darstellte, erhöht wurden.

Beim Estes Modell werden in einem Fenster, das vier Wörter enthält, die Assoziationsstärken von sechs Wortpaaren in einem Lerndurchgang erhöht, während beim Rescorla & Wagner Modell nur bei drei Wortpaaren die Assoziationsstärken erhöht werden.

Die folgenden Simulationen sollen klären, ob die Verstärkung der geringeren Anzahl von Wortpaaren die Ursache für das schlechtere Abschneiden des Rescorla & Wagner Modells bei der Berechnung von Wortassoziationen war. Um dies zu testen, wurde die Anzahl der zu verstärkenden Wortpaare pro Lerndurchgang erhöht. Die Regel für Vergessen wurde bei diesem Test nicht mehr ausgeführt, da sie im vorherigen Abschnitt nicht zu verbesserten Ergebnissen geführt hatte. Ein weiterer Grund, der zum Weglassen der Regel für Vergessen geführt hatte, war der zu erwartende starke Anstieg der Rechenzeit.

Die Anzahl der zu verstärkenden Wortpaare wurde erhöht, indem jedes Wort im Fenster einmal als US betrachtet wurde und die jeweils anderen Worte im Fenster die Reize darstellten. Dies wurde durch eine Rotation der Worte im Fenster realisiert. Jedes Wort im Fenster stand durch die Rotation einmal als letztes Wort im Fenster und wurde somit ein US.

Faktor F	Pot. L	Pot. V	K 1	K 2	K 3	K 4	Fensterlänge
$3.6E - 18$	6	5	27	287	40	52	22
$6E - 18$	6	5	28	282	39	51	22
$9E - 18$	6	5	28	282	38	50	22
$3.6E - 19$	6	5	26	287	39	50	22
$3.6E - 18$	6	4	26	288	39	50	22
$3.6E - 15$	5	4	26	272	38	53	22
$3.6E - 16$	5	4	26	285	36	51	22
$3.6E - 17$	5	4	26	287	36	51	22
$3.6E - 18$	5	4	26	287	36	51	22

Tabelle 6.18:

Rescorla & Wagner Modell: Ergebnisse, die mit den unterschiedlichen Parametern für Lernen und Vergessen erzielt wurden.

$$\text{Lernparameter: } \alpha_i \beta_j = F \cdot (I_i \cdot I_j)^L$$

$$\text{Parameter für Vergessen: } \alpha_i \beta_j = F \cdot (I_i \cdot I_j)^V$$

In der ersten Spalte ist der Faktor angegeben, der bei der Ausführung der Lernregel und der Regel für Vergessen verwendet wurde. Die zweite Spalte enthält die Potenzierung für die Lernregel, während die dritte Spalte die Potenzierung bei Anwendung der Regel für Vergessen enthält. Die nachfolgenden Spalten enthalten die Ergebnisse in Bezug auf die vier Kriterien und die zugehörige Fensterlänge.

An einem Fenster, das vier Worte enthält, wird die Rotation erklärt:

$$\boxed{w_1 \ w_2 \ w_3 \ \boxed{w_4}}$$

Das Fenster enthält die Wörter w_1 , w_2 , w_3 und w_4 . Im ersten Schritt ist das Wort w_4 der US, die anderen Wörter w_1 , w_2 und w_3 sind die Reize. Es werden dann die Wortpaare (w_1, w_4) , (w_2, w_4) und (w_3, w_4) verstärkt. Ohne Rotation würde das Fenster jetzt um ein Wort nach rechts verschoben. Da die Wörter im Fenster aber erst noch rotiert werden sollen, wird das Fenster noch nicht nach rechts verschoben.

Alle Worte im Fenster verändern ihre Position, indem sie eine Position nach rechts geschoben werden. Das Wort w_3 wird dadurch zum US. Das Wort w_4 gleitet aus dem Fenster heraus und wird vorne an das Fenster wieder angehängt.

$$\boxed{w_4 \ w_1 \ w_2 \ \boxed{w_3}}$$

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$3.6E - 12$	25	262	35	51	22
$3.6E - 13$	25	254	35	49	22
$3.6E - 14$	25	253	35	50	22
$3.6E - 15$	25	253	35	50	22

Tabelle 6.19: Rescorla & Wagner Modell ohne Regel für Vergessen: Ergebnisse, die mit dem Lernparameter $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^4$ erzielt wurden. Die dargestellten Ergebnisse in Bezug auf die vier Kriterien unterscheiden sich bei unterschiedlichen Faktoren kaum voneinander. Bei der Potenzierung mit vier wurden keine guten Ergebnisse in Bezug auf das zweite Kriterium erzielt.

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$3.6E - 15$	26	286	38	52	24
$3.6E - 16$	26	288	36	51	22
$6.6E - 16$	26	288	36	51	22
$3.6E - 17$	26	288	36	51	22

Tabelle 6.20: Rescorla & Wagner Modell ohne Regel für Vergessen: Ergebnisse, die mit dem Lernparameter $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^5$ erzielt wurden. Die dargestellten Ergebnisse in Bezug auf die vier Kriterien unterscheiden sich bei unterschiedlichen Faktoren kaum voneinander. Durch die Potenzierung mit 5 konnten die Ergebnisse hauptsächlich in Bezug auf das zweite Kriterium verbessert werden.

Nach der ersten Rotation werden die folgenden Wortpaare (w_4, w_3) , (w_1, w_3) und (w_2, w_3) verstärkt. Danach erfolgt die zweite Rotation, d.h. das Wort w_2 wird zum US.

$$\boxed{w_3 \ w_4 \ w_1 \ \boxed{w_2}}$$

Nach der zweiten Rotation werden die folgenden Wortpaare (w_3, w_2) , (w_4, w_2) und (w_1, w_2) verstärkt. Danach erfolgt noch eine weitere Rotation. Durch die letzte Rotation dieses Fensters wird das Wort w_1 zum US.

$$\boxed{w_2 \ w_3 \ w_4 \ \boxed{w_1}}$$

Nach der letzten Rotation werden die Wortpaare (w_2, w_1) , (w_3, w_1) und (w_4, w_1) verstärkt.

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$1E - 17$	26	286	38	50	22
$3.6E - 18$	26	288	39	50	22
$8.6E - 18$	26	287	39	50	22
$3.6E - 19$	26	286	39	50	22
$3.6E - 20$	26	286	39	50	22

Tabelle 6.21: Rescorla & Wagner Modell ohne Regel für Vergessen: Ergebnisse, die mit dem Lernparameter $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^6$ erzielt wurden. Die dargestellten Ergebnisse in Bezug auf die vier Kriterien unterscheiden sich bei unterschiedlichen Faktoren kaum voneinander.

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$3.6E - 16$	28	287	38	50	22
$3.6E - 17$	28	287	38	50	22

Tabelle 6.22: Rescorla & Wagner Modell ohne Regel für Vergessen: Ergebnisse, die mit dem Lernparameter $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^{5.5}$ erzielt wurden. Die dargestellten Ergebnisse in Bezug auf die vier Kriterien unterscheiden sich bei unterschiedlichen Faktoren kaum voneinander.

Durch die Rotation wird jedes Wort im Fenster genau einmal zum US, während die jeweils anderen Wörter die zugehörigen Reize darstellen. In einem Lerndurchgang werden durch die Rotation 12 Wortpaare verstärkt. Ohne Rotation werden nur drei Wortpaare verstärkt. Das führt dazu, dass jetzt doppelt so viele Wortpaare wie beim Estes Modell verstärkt werden. Beim Estes Modell werden bei einem Fenster, das vier Worte enthält, sechs Wortpaare verstärkt. Dies liegt daran, dass das Rescorla & Wagner Modell nicht symmetrisch ist. Allgemein werden bei einem Fenster, das aus n Wörtern besteht, $n(n - 1)$ Wortpaare pro Lerndurchgang verstärkt, wenn der Fensterinhalt in jedem Lerndurchgang nach der obigen Beschreibung rotiert wird.

In den Tabellen 6.23 und 6.24 werden die Ergebnisse dargestellt, die durch die Rotation des Fensterinhaltes erzielt wurden.

Durch die Rotation des Fensters verbesserten sich die Ergebnisse in Bezug auf das erste Kriterium. Es konnten jetzt 30 Primärantworten richtig vorausgesagt werden. Die Anzahl aller Responsewörter der Teilnehmer, die auf den ersten zehn Rangplätzen gezählt wurden, erhöhte sich dabei auf 300 (bei gleichzeitig 30 richtig vorausgesagten Primärantworten) bzw. 311 (bei 29 richtig vorausgesagten Primärantworten). Diese durch die Rotation des Fensters erzielten Ergebnisse waren jetzt genauso gut wie die Ergebnis-

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$3.6E - 19$	29	311	39	56	36
$5E - 19$	29	311	39	56	36
$1E - 20$	30	299	38	54	62
$3.6E - 20$	29	306	38	55	22
$3.6E - 20$	30	299	38	54	62
$3.6E - 21$	29	306	38	54	22
$3.6E - 21$	30	300	38	54	62
$3.6E - 22$	29	306	38	54	22
$3.6E - 22$	30	300	38	54	62

Tabelle 6.23: Rescorla & Wagner Modell: Ergebnisse, die mit dem Lernparameter $\alpha_i\beta_j = F \cdot (I_i \cdot I_j)^5$ und einer Rotation des Fensterinhaltes erzielt wurden. Die aufgeführten Ergebnisse unterscheiden sich bei den unterschiedlichen Faktoren in Bezug auf die vier Kriterien nur wenig voneinander. Mit den Faktoren $3.6E - 20$, $3.6E - 21$ und $3.6E - 22$ wurde ein gutes Ergebnis in Bezug auf alle vier Kriterien schon bei einer Fensterlänge von 22 erreicht.

se, die mit dem Estes Modell erzielt wurden. Die maximale Anzahl richtig vorausgesagter Primärantworten wurde aber erst bei einer Fensterlänge von 62 erzielt. Dies wurde auch schon beim Estes Modell beobachtet. Bei einer Fensterlänge von 22 konnten mit den Parametern, die die meisten richtigen Primärantworten voraussagen, aber auch schon bessere oder gleich gute Ergebnisse in Bezug auf die anderen drei Kriterien erzielt werden.

Eine Potenzierung mit 6 ergab eine leichte Verschlechterung der Ergebnisse in Bezug auf das erste Kriterium, es wurden 2 Primärantworten weniger vorausgesagt. In Bezug auf das zweite Kriterium verbesserten sich die Ergebnisse etwas gegenüber den Ergebnissen, die mit einer Potenzierung von 5 erzielt wurden. Dieser Effekt wurde auch bei den Simulationen mit dem Estes Modell beobachtet.

Für eine Simulationsreihe wurde die Regel für Vergessen wieder eingeführt und der Fensterinhalt rotiert. Diese Simulationsreihe konnte wegen der langen Rechenzeit nur bis zu einer Fensterlänge von 50 durchgeführt werden. Der Faktor in der Regel für Vergessen wurde eine zehner Potenz kleiner als der Faktor in der Regel für Lernen gewählt. Als Potenzierung wurde $n = 5$ gewählt, da bei dieser Potenzierung schon ohne Regel für Vergessen die besten Resultate erzielt wurden. Die Parameter für Lernen und Vergessen werden durch folgende Formeln spezifiziert:

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$3.6E - 18$	27	304	40	52	22
$3.6E - 19$	27	321	40	53	42
$5E - 19$	28	321	39	53	42
$3.6E - 20$	28	316	40	52	84
$5E - 20$	28	317	39	52	82
$3.6E - 21$	28	316	39	54	90

Tabelle 6.24: Rescorla & Wagner Modell: Ergebnisse, die mit dem Lernparameter $\alpha_i \beta_j = F \cdot (I_i \cdot I_j)^6$ und einer Rotation des Fensterinhaltes erzielt wurden. Bei einem Faktor von $5E - 19$ wurde das beste Ergebnis in Bezug auf die vier Kriterien erzielt. Dieses Ergebnis konnte aber erst mit einer Fensterlänge von 42 erzielt werden.

Lernen:

$$\alpha_i \cdot \beta_j = 3.6E - 20 \cdot (I_i \cdot I_j)^5 \quad (6.9)$$

Vergessen:

$$\alpha_i \cdot \beta_j = 3.6E - 21 \cdot (I_i \cdot I_j)^5 \quad (6.10)$$

Die erzielten Ergebnisse in Bezug auf die vier Kriterien unterschieden sich nur unwesentlich von den Ergebnissen, die ohne Anwendung der Regel für Vergessen erzielt wurden.

Die Ergebnisse in Bezug auf das erste Kriterium in Abhängigkeit von der Fensterlänge sind in der Abbildung 6.3 dargestellt. Die helle Kurve mit den Rechtecken gehört zu den Ergebnissen, die ohne die Rotation des Fensterinhaltes erzielt wurden, aber mit Anwendung der Regel für Vergessen. Die dunkle Kurve mit den Rauten repräsentiert die Ergebnisse, die mit der Rotation des Fensterinhaltes erzielt wurden. Die Ergebnisse, die zur hellen Kurve gehören, wurden mit einer Potenzierung von 6 erzielt. Für die Lernregel wurde der Faktor $6E - 18$ und für die Regel für Vergessen der Faktor $6E - 20$ gewählt. Die helle Kurve in Abbildung 6.3 steht repräsentativ für alle Parameter, bei denen 28 richtig vorausgesagte Primärantworten berechnet wurden. Bei den anderen Parametern, die gleich gute Ergebnisse, d.h. 28 richtig vorausgesagte Primärantworten, lieferten, unterschieden sich die Ergebnisse nur wenig von denen, die in der Abbildung 6.3 dargestellt sind.

Die Ergebnisse in Bezug auf das zweite Kriterium in Abhängigkeit von der Fensterlänge sind in der Abbildung 6.4 dargestellt. Die helle Kurve gehört zu den Ergebnissen, die ohne Rotation des Fensterinhaltes erzielt wurden. Die

dunkle Kurve gehört zu den Ergebnissen, die mit einer Rotation des Fensterinhaltes berechnet wurden. Eine Rotation des Fensterinhaltes verbesserte die Ergebnisse in Bezug auf das zweite Kriterium (siehe Abbildung 6.4). Die besten Ergebnisse in Bezug auf das zweite Kriterium wurden nicht bei einer Fensterlänge von 22 erreicht, wie bei Kriterium 1, sondern bei einer Fensterlänge von 32 (helle Kurve: 287 Nennungen) und bei einer Fensterlänge von 36 (dunkle Kurve: 309 Nennungen). Bei einer Fensterlänge von 22 wurden aber nur leicht schlechtere Ergebnisse in Bezug auf das zweite Kriterium erzielt (helle Kurve: 283 Nennungen, dunkle Kurve: 306 Nennungen).

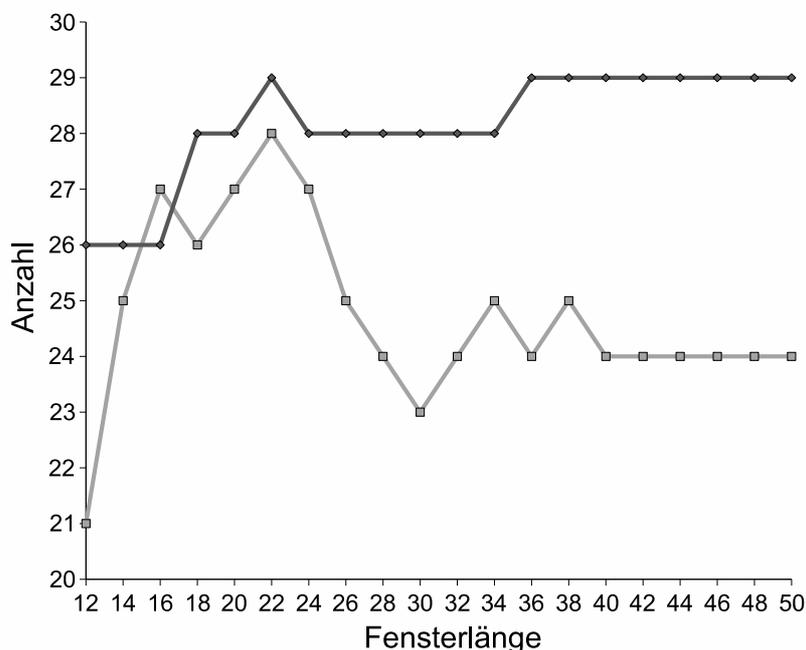


Abbildung 6.3: Die Anzahl der richtig vorhergesagten Primärantworten (Kriterium 1), die mit dem Rescorla & Wagner Modell gewonnen wurden, in Abhängigkeit von der Fensterlänge.

Die helle Kurve repräsentiert die Ergebnisse, die ohne Rotation des Fensterinhaltes entstanden waren. Die dunkle Kurve entstand, wenn der Fensterinhalt rotiert wurde. Ohne Rotation des Fensters (helle Kurve mit Rechtecken) wurden bei einer Fensterlänge von 22 die meisten Primärantworten (28) richtig vorausgesagt. Wenn das Fenster rotiert wurde (dunkle Kurve mit Rauten), wurden bei einer Fensterlänge von 22 ebenfalls die meisten Primärantworten (29) richtig vorausgesagt. Bei den Fensterlängen von 24 bis 34 wurde eine Primärantwort weniger vorausgesagt als bei einer Fensterlänge von 22. Ab einer Fensterlänge von 36 konnten wieder 29 Primärantworten richtig vorausgesagt werden. Ohne Rotation des Fensterinhaltes wurden nur bei einer Fensterlänge von 22 die meisten Primärantworten richtig vorausgesagt. Ab einer Fensterlänge von 24 sank die Anzahl der richtig vorausgesagten Primärantworten ab, so dass ab einer Fensterlänge von 40 nur 24 Primärantworten vorausgesagt werden konnten. Durch die Rotation des Fensterinhaltes änderte sich ab einer Fensterlänge von 18 die Anzahl der richtig vorausgesagten Primärantworten kaum. Sie lag bei 28(+1).

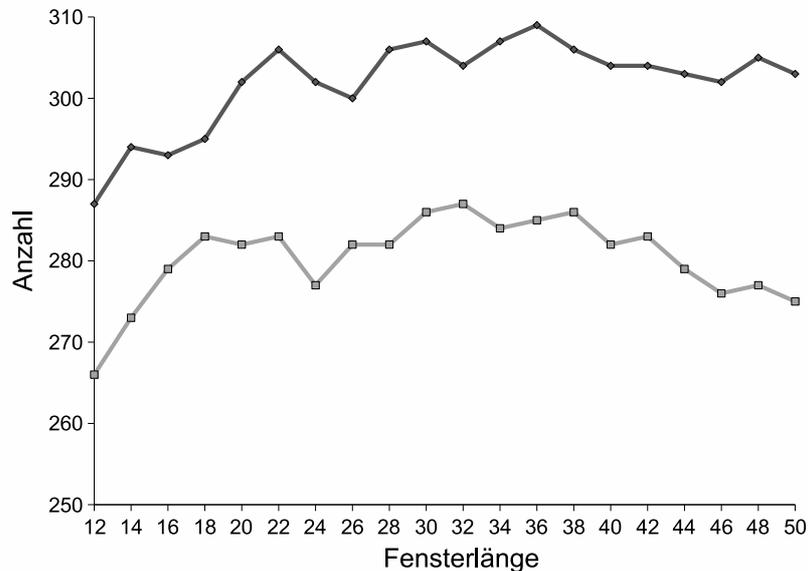


Abbildung 6.4: Die Ergebnisse, die mit dem Rescorla & Wagner Modell, in Bezug auf das Kriterium 2 in Abhängigkeit von der Fensterlänge erzielt wurden.

Die untere Kurve mit den Rechtecken entstand, wenn der Fensterinhalt nicht rotiert wurde. Mit Rotation des Fensterinhaltes wurden die Ergebnisse erzielt, die durch die obere Kurve dargestellt sind. Bei einer Rotation des Fensterinhaltes wurde bei einer Fensterlänge von 36 das beste Ergebnis in Bezug auf das zweite Kriterium erzielt. Dies waren 309 richtig vorausgesagte Nennungen der Teilnehmer ohne die Primärantworten auf den ersten zehn Rangplätzen. Ohne Rotation des Fensterinhaltes wurde bei einer Fensterlänge von 32 das beste Ergebnis in Bezug auf das zweite Kriterium erzielt. Dies waren 287 richtig vorausgesagte Nennungen der Teilnehmer auf den ersten zehn Rangplätzen. Ab einer Fensterlänge von 20 wurden mindestens 300 Nennungen auf den ersten zehn Rangplätzen richtig vorausgesagt, wenn der Fensterinhalt rotiert wurde.

Für die Simulationen mit den Testkorpora wurden drei Parameter ausgewählt. Beim ersten Parameter wurde sowohl Lernen als auch Vergessen berücksichtigt. Der Lernparameter unterschied sich vom Parameter für Vergessen nur durch einen unterschiedlichen Faktor, während sich die Potenzierung nicht änderte (Potenzierung $n = 6$). Der Faktor für die Lernregel hatte den Wert $3.6E - 18$ und beim Vergessen wurde ein Faktor von $3.6E - 20$ verwendet. Beim zweiten und dritten Parameter wurde die Regel für Vergessen nicht mehr berücksichtigt, da diese nicht zu besseren Ergebnissen geführt

hatte. Der zweite Lernparameter bestand aus einem Faktor von $3.6E - 16$ und einer Potenzierung von 5.5, da mit diesem Parameter die besten Ergebnisse in Bezug auf die vier Kriterien ermittelt wurden. Bei der Verwendung des dritten Lernparameters wurde zusätzlich noch der Fensterinhalt rotiert. Die verwendete Potenzierung lag bei $n = 5$, während der zugehörige Faktor einen Wert von $3.6E - 21$ hatte.

Alle drei Parameter lieferten gute Ergebnisse in Bezug auf die vier Kriterien mit einer Fensterlänge von 22.

6.4.5 Gallistel Modell

Das Gallistel Modell besitzt keine freien Parameter, die bestimmt werden mussten. Bei der Simulation von Wortassoziationen war aber die Fensterlänge ein freier Parameter, der bestimmt werden musste.

Wurde das Gallistel Modell mit unterschiedlichen Fensterlängen simuliert, so fiel ein starker Worthäufigkeitseffekt bei den berechneten Assoziationen auf. Dieser starke Worthäufigkeitseffekt war beim Rescorla & Wagner Modell bei der Verwendung eines konstanten Parameters beobachtet worden. Sehr häufig wurde als Primärantwort das häufigste Wort im Vokabular (sein) berechnet. Die von den Teilnehmern genannte Primärantwort war aber häufig schon auf dem zweiten Rangplatz zu finden. Erst ab einer Fensterlänge von 64 konnten mehr als 10 richtig vorausgesagte Primärantworten vom Gallistel Modell auf dem ersten Rangplatz berechnet werden. Bei einer Fensterlänge von 74 wurden 17 richtig vorausgesagte Primärantworten auf dem ersten Rangplatz gezählt. Dies war das beste Ergebnis in Bezug auf das erste Kriterium. Der Worthäufigkeitseffekt war auch bei dieser Fensterlänge noch deutlich zu erkennen.

Um den Worthäufigkeitseffekt zu verringern, wurde der erste Lösungsansatz zur Verringerung des Worthäufigkeitseffekts aus dem Abschnitt Worthäufigkeitseffekt verwendet. Der zweite Lösungsansatz zur Verringerung des Worthäufigkeitseffekts konnte nicht verwendet werden, da das Gallistel Modell keinen Lernparameter enthält.

Beim ersten Lösungsansatz zur Verringerung des Worthäufigkeitseffekts wurde die berechnete Assoziationsstärke mit dem Informationsgehalt des jeweiligen Responsewortes gewichtet. Durch diese Gewichtung konnten die berechneten Assoziationen verbessert werden, da der Worthäufigkeitseffekt verringert wurde. Jedoch waren weiterhin häufige Wörter vermehrt auf den ersten zehn Rangplätzen zu beobachten. Um den Worthäufigkeitseffekt noch weiter zu verringern, wurde bei der Gewichtung der Informationsgehalt des Responsewortes I_j vor der Multiplikation mit der Assoziationsstärke $a_{i,j}$ noch potenziert.

$$g_{i,j} = a_{i,j} \cdot I_j^n \quad (6.11)$$

Durch die Potenzierung mit n erhielt das Gallistel Modell zur Simulation von Wortassoziationen einen weiteren freien Parameter, der bestimmt werden musste. Zur Bestimmung von n in der Gleichung 6.11 wurde die Potenz n schrittweise um eins erhöht. Danach wurden die erzielten Ergebnisse in Bezug auf die vier Kriterien bewertet. Die Tabelle 6.25 zeigt die Ergebnisse in Bezug auf die vier Kriterien, die durch die unterschiedlichen Potenzierungen entstanden waren, in tabellarischer Form.

n	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
0	17	138	18	30	74
1	20	187	25	39	96
2	22	215	32	43	96
3	24	243	30	51	18
4	26	271	34	52	24
5	27	277	36	52	22
6	26	279	34	47	32

Tabelle 6.25: Gallistel Modell: Ergebnisse der Simulationen, wenn nach der Berechnung noch eine Gewichtung vorgenommen wurde. Die Gewichtung wurde mit der Formel $g_{i,j} = a_{i,j} \cdot I_j^n$ vorgenommen. In der ersten Spalte ist die Potenzierung des Informationsgehaltes des Responsewortes aufgeführt. Die erste Zeile enthält die Ergebnisse in Bezug auf die vier Kriterien, die ohne Gewichtung erzielt wurden. Es wurden Potenzierungen bis $n = 6$ durchgeführt. Ab einer Potenzierung mit $n = 4$ verbesserten sich die Ergebnisse in Bezug auf die vier Kriterien. Mit einer Potenzierung von 5 wurde bei einer Fensterlänge von 22 das beste Ergebnis in Bezug auf alle vier Kriterien erzielt.

Die erste Zeile der Tabelle 6.25 enthält die Ergebnisse in Bezug auf die vier Kriterien, die ohne Gewichtung erzielt wurden. Wurde nur mit dem Informationsgehalt des Responsewortes gewichtet (keine Potenzierung, $n = 1$), so wurde erst bei einer Fensterlänge von 96 die maximale Anzahl richtig vorausgesagter Primärantworten erreicht.

Erst ab einer Potenzierung mit $n = 5$ konnten die besten Ergebnisse in Bezug auf alle vier Kriterien erzielt werden. Weitere Simulationen, bei denen Potenzierungen im Intervall von 4.2 bis 6.2 durchgeführt wurden, werden in der Tabelle 6.26 dargestellt.

Die besten Ergebnisse in Bezug auf alle vier Kriterien wurden mit Potenzierungen im Intervall von 5.4 bis 5.8 erzielt.

Der Worthäufigkeitseffekt konnte deutlich verringert werden, wenn die Assoziationsstärke mit dem potenzierten Informationsgehalt des Responsewortes gewichtet wurde. Es folgen zwei Beispiele, bei denen einmal die ungewichteten Assoziationen, siehe hierzu die Tabelle 6.27 und die Tabelle 6.29, und einmal die gewichteten Assoziationen, siehe hierzu die Tabelle 6.28 und die Tabelle 6.30, gezeigt werden.

Nach einer Gewichtung der Responsewörter zu dem Stimuluswort *essen* mit der Formel 6.11, wobei n den Wert 5.6 hatte, ergaben sich die Assoziationen, die in der Tabelle 6.28 aufgelistet sind. Die Assoziationsstärken

n	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
4.2	26	275	34	52	24
4.5	26	275	35	52	24
4.6	26	277	35	51	24
4.7	26	277	36	52	24
4.8	26	279	37	52	24
4.9	27	279	36	51	22
5.1	27	279	36	52	22
5.2	27	278	37	52	24
5.3	27	276	38	52	24
5.4	28	275	37	51	22
5.5	28	272	37	50	22
5.6	28	277	37	50	22
5.7	28	275	39	50	24
5.8	28	276	39	50	24
5.9	27	278	38	48	22
6.2	24	277	35	46	34

Tabelle 6.26: Gallistel Modell: Ergebnisse der Simulationen, wenn nach der Berechnung noch eine Gewichtung vorgenommen wurde. Die Gewichtung wurde mit der Formel $g_{i,j} = a_{i,j} \cdot I_j^n$ vorgenommen. Es werden die Potenzierungen zwischen 4.2 und 6.2 in Bezug auf die vier Kriterien dargestellt. Die besten Ergebnisse in Bezug auf das erste Kriterium wurden bei den Potenzierungen von 5.4 bis 5.8 erzielt. Dagegen konnten die besten Ergebnisse beim zweiten Kriterium schon ab einer Potenzierung von 4.2 erzielt werden. Bei allen Potenzierungen im Intervall von 4.2 bis 6.2 wurden in Bezug auf das zweite Kriterium gute Ergebnisse erzielt, die bei den unterschiedlichen Potenzierungen nur sehr wenig voneinander abwichen.

der Responsewörter wurden nach der Gewichtung noch mit der euklidischen Norm normiert, um wieder Assoziationsstärken zwischen 0 und 1 zu erhalten.

Die häufigen Wörter, die ohne Gewichtung zu dem Stimuluswort *essen* assoziiert wurden, waren jetzt nicht mehr auf den ersten fünfzehn Rangplätzen zu finden.

Auch bei den berechneten Assoziationen zu *Kind*, die in der Tabelle 6.29 aufgeführt werden, wurden viele häufige Wörter aus dem Vokabular zum Stimuluswort als assoziative Antworten berechnet. Durch die Gewichtung konnten diese häufigen Wörter von den ersten fünfzehn Rangplätzen verdrängt werden. Die Assoziationen zu *Kind* konnten hierdurch deutlich verbessert werden (siehe hierzu auch Tabelle 6.30).

Rangplatz	Assoziationsstärke	Responsewort
1	0.8857	sein
2	0.3269	trinken
3	0.1387	wollen
4	0.1268	gehen
5	0.0854	Kind
6	0.0779	denken
7	0.0644	lang
8	0.0612	Frau
9	0.0556	schlafen
10	0.0555	klein
11	0.0487	Mahlzeit
12	0.0474	Brot
13	0.0414	Fleisch
14	0.0390	Tisch
15	0.0364	Mensch

Tabelle 6.27: Ungewichtete Assoziationen zu dem Stimuluswort *essen*, die mit dem Gallistel Modell vorausgesagt wurden.

Das häufigste Wort *sein* im Vokabular hatte eine auffallend hohe Assoziationsstärke zum Stimuluswort. Unter den ersten 15 Responsewörtern zu dem Stimuluswort *essen* waren viele häufige Wörter aus dem Vokabular (*sein*, *wollen*, *gehen*, *lang*, *klein*, *Kind*, *Frau* und *Mensch*) vertreten.

Wenn mit dem Gallistel Modell Wortassoziationen berechnet werden sollten, so mussten die Assoziationsstärken nach der Berechnung noch gewichtet werden, damit die Ergebnisse nicht zu stark durch den Worthäufigkeitseffekt verfälscht wurden. Der Vorteil des Gallistel Modells gegenüber den Estes Modell und dem Rescorla & Wagner Modell ist ja, dass es keine freien Parameter enthält, die bestimmt werden müssen (außer der Fensterlänge). Für die Berechnung von freien Wortassoziationen wurde aber ein freier Parameter benötigt, da ohne freien Parameter die Ergebnisse zu stark durch den Worthäufigkeitseffekt verfälscht wurden.

Im folgenden Abschnitt werden die Ergebnisse, die mit dem Gallistel Modell erzielt wurden, mit den Ergebnissen des Rescorla & Wagner Modells ohne Rotation verglichen. Die Simulationen des Rescorla & Wagner Modells, die keine Rotation des Fensterinhaltes verwendet hatten, und das Gallistel Modell mit Gewichtung unterschieden sich in den Ergebnissen in Bezug auf die vier Kriterien bei unterschiedlichen Fensterlängen nur wenig voneinander. In der Abbildung 6.5 werden die Ergebnisse in Bezug auf das erste Kriteri-

Rangplatz	Assoziationsstärke	Responsewort
1	0.714	trinken
2	0.295	Mahlzeit
3	0.206	Blumenkohl
4	0.156	Brot
5	0.141	Schokolade
6	0.132	Fleisch
7	0.128	Reis
8	0.127	Suppe
9	0.119	Hunger
10	0.118	Salat
11	0.115	schlafen
12	0.106	Teller
13	0.105	Butterbrot
14	0.097	Speise
15	0.081	Fett

Tabelle 6.28: Gewichtete Assoziationen zu dem Stimuluswort *essen*, die mit dem Gallistel Modell erzeugt wurden. Die Gewichtung erfolgte mit der Formel $g_{i,j} = a_{i,j} \cdot I_j^{5.6}$

Die häufigen Wörter, die ohne Gewichtung zu dem Stimuluswort *essen* assoziiert wurden, waren jetzt nicht mehr auf den ersten fünfzehn Rangplätzen zu finden.

um in Abhängigkeit von der Fensterlänge dargestellt. Die helle Kurve mit den Rechtecken repräsentiert das Rescorla & Wagner Modell ohne Rotation und die dunkle Kurve mit den Rauten das Gallistel Modell mit Gewichtung (Gewichtung mit einer Potenzierung von $n = 5.6$).

In der Abbildung 6.6 werden die Ergebnisse in Bezug auf das zweite Kriterium in Abhängigkeit von der Fensterlänge dargestellt. Von beiden Modellen konnten 287 Nennungen der Teilnehmer ohne die Primärantworten auf den ersten zehn Rangplätzen mit einer Fensterlänge von 32 richtig vorausgesagt werden. Bei dieser Fensterlänge wurden aber nur 24 Primärantworten vom Rescorla & Wagner Modell und 25 Primärantworten vom Gallistel Modell richtig auf den ersten Rangplatz vorausgesagt. Bei einer Fensterlänge von 22 wurden vom Rescorla & Wagner Modell 283 Nennungen auf den ersten zehn Rangplätzen richtig vorausgesagt, während das Gallistel Modell bei dieser Fensterlänge 11 Nennungen weniger voraussagte.

Mit einer Gewichtung lieferte das Gallistel Modell bei einer Fensterlänge von 22 gute Ergebnisse in Bezug auf alle vier Kriterien. Für die Simulatio-

Rangplatz	Assoziationsstärke	Responsewort
1	0.9362	sein
2	0.2236	Eltern
3	0.1074	Mutter
4	0.0785	klein
5	0.0740	Frau
6	0.0620	Seite
7	0.0533	wollen
8	0.0487	Konzert
9	0.0411	Welt
10	0.0410	Haus
11	0.0379	Straße
12	0.0356	gehen
13	0.0350	lernen
14	0.0331	Mädchen
15	0.0261	jung

Tabelle 6.29: Ungewichtete Assoziationen zu dem Stimuluswort *Kind*, die mit dem Gallistel Modell vorausgesagt wurden. Viele häufige Wörter aus dem Vokabular sind auf den ersten fünfzehn Rangplätzen.

nen mit den Testkorpora wurden zwei unterschiedliche Gewichtungen ausgewählt, einmal eine Potenzierung mit $n = 5.6$ und einmal eine Potenzierung mit $n = 5$. Bei einer Fensterlänge von 22 lieferte eine Potenzierung mit $n = 5.6$ die besten Ergebnisse in Bezug auf alle vier Kriterien. Da sich ab einer Potenzierung von $n = 5$ die Ergebnisse in Bezug auf die vier Kriterien kaum voneinander unterschieden hatten, wurde für die zweite Gewichtung eine Potenzierung von $n = 5$ gewählt.

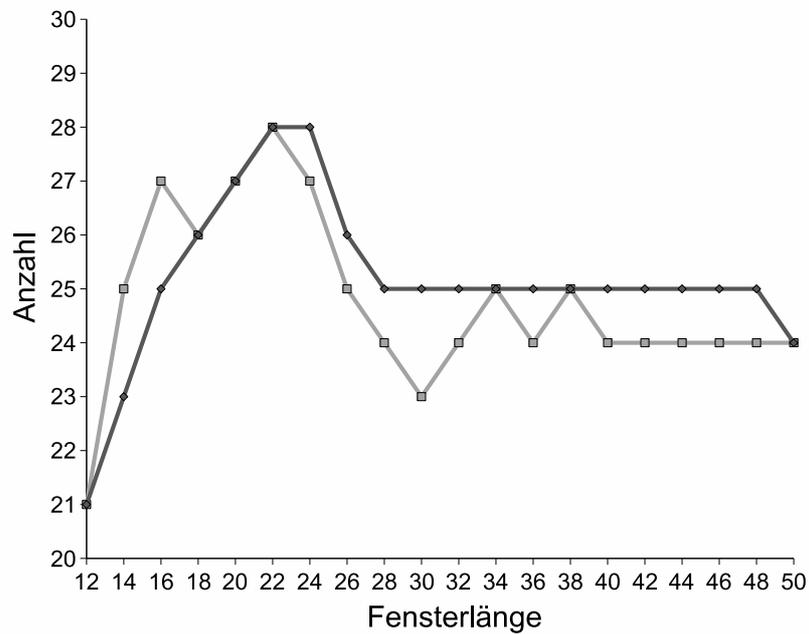


Abbildung 6.5: Modellvergleich Gallistel Modell und Rescorla & Wagner Modell in Bezug auf die Anzahl Primärantworten (Kriterium 1) in Abhängigkeit von der Fensterlänge.

Die helle Kurve gehört zu den Ergebnissen, die mit dem Rescorla & Wagner Modell produziert wurden. Der Fensterinhalt wurde dabei nicht rotiert. Die dunkle Kurve gehört zu den Ergebnissen, die mit dem Gallistel Modell erzeugt wurden. Beide Modelle konnten die meisten Primärantworten (28) bei einer Fensterlänge von 22 voraussagen.

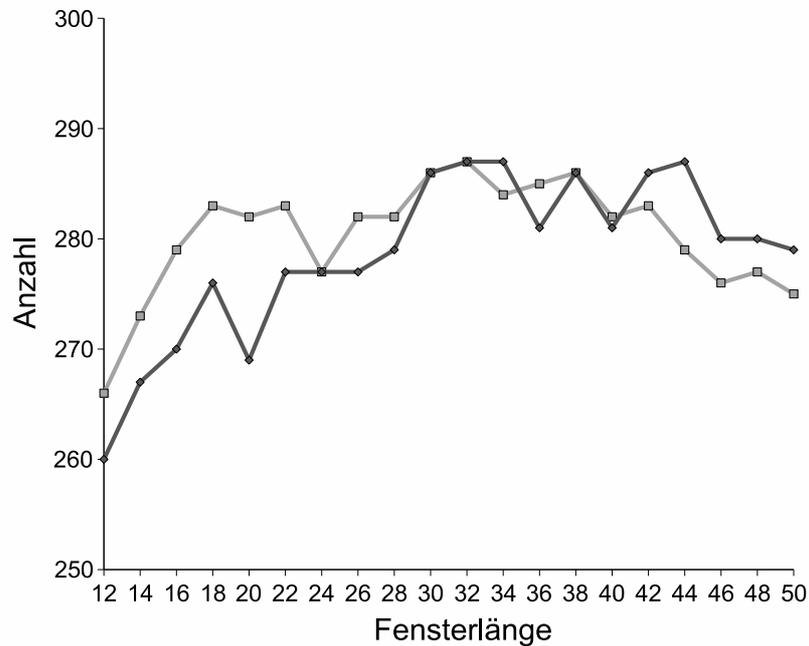


Abbildung 6.6: Modellvergleich Gallistel Modell und Rescorla & Wagner Modell in Bezug auf das Kriterium 2 in Abhängigkeit von der Fensterlänge: Die Ergebnisse, die die helle Kurve mit den Rechtecken repräsentiert, wurden mit dem Rescorla & Wagner Modell ohne Rotation produziert. Die Ergebnisse, die die dunkle Kurve repräsentiert, wurden mit dem Gallistel Modell erzeugt. Das Gallistel Modell erzielte die besten Ergebnisse in Bezug auf das zweite Kriterium bei den Fensterlängen von 32, 34 und 44 (287 Nennungen). Das Rescorla & Wagner Modell erzielte bei einer Fensterlänge von 32 die besten Ergebnisse in Bezug auf das zweite Kriterium (ebenfalls 287 Nennungen).

Rangplatz	Assoziationsstärke	Responsewort
1	0.7003	Eltern
2	0.3366	Mutter
3	0.1997	Konzert
4	0.1259	Mädchen
5	0.1149	Bub
6	0.1136	Baby
7	0.1127	Kegel
8	0.112	Schule
9	0.1107	lernen
10	0.1005	Heim
11	0.09918	klein
12	0.09052	Babysitter
13	0.08928	Lehrer
14	0.08284	krank
15	0.0828	Junge

Tabelle 6.30: Gewichtete Assoziationen zu dem Stimuluswort *Kind*, die mit dem Gallistel Modell erzeugt wurden. Die Gewichtung erfolgte mit der Formel $g_{i,j} = a_{i,j} \cdot I_j^{5.6}$.

Der starke Worthäufigkeitseffekt, der bei den ungewichteten Assoziationen zu dem Stimuluswort *Kind* beobachtet wurde, ist nicht mehr vorhanden. Die starke Assoziation von *Konzert* zu *Kind* entstand durch das häufige gemeinsame Auftreten von *Kind* mit *Konzert* in den Veranstaltungshinweisen der Zeitungen. Bei den Konzertankündigungen liest man häufig, dass Konzerte im Kinder- und Jugendmuseum stattfinden. In dem lemmatisierten Textkorpus SZ98 sieht eine solche Konzertankündigung folgendermaßen aus:

Kind . und Jugendmuseum . Arnulfstr . 3 Konzert Kammermusikabend.

Der Bindestrich fiel beim *Kind* weg und wurde durch einen Punkt ersetzt, da in der Vorverarbeitung des Textkorpus alle Sonderzeichen durch Punkte ersetzt wurden.

Kapitel 7

Ergebnisse

In dem vorherigen Kapitel wurden die Werte für die freien Parameter, die in den einzelnen Lernmodellen enthalten sind, anhand eines Kalibrierkorpus bestimmt. In diesem Kapitel soll überprüft werden, ob die mit dem Kalibrierkorpus ermittelten freien Parameter auch bei Verwendung der Testkorpora gute Ergebnisse in Bezug auf die vier Kriterien liefern werden. Erst dann darf auf die Allgemeingültigkeit der Ergebnisse geschlossen werden.

Für jedes Lernmodell wurden Simulationen mit den vier verschiedenen Zeitungskorpora durchgeführt (SZ95, SZ96, SZ97 und die gesamte SZ). Dies sind die Testkorpora, die im vorherigen Kapitel im Abschnitt Korpora beschrieben wurden.

Alle Lernmodelle hatten mit einer Fensterlänge von 22 gute Ergebnisse in Bezug auf die vier Kriterien erzielt, wenn das Kalibrierkorpus verwendet wurde. Deswegen wurden die Simulationen mit den Testkorpora ebenfalls mit einer Fensterlänge von 22 durchgeführt.

In dem vorherigen Kapitel wurden im Abschnitt Parameterschätzung unterschiedliche Parameter für die Simulationen mit den Testkorpora ausgewählt. Diese werden im folgenden Abschnitt kurz zusammengefasst.

Für das Estes Modell wurden zwei unterschiedliche Parameter ausgewählt, einmal ein konstanter Parameter für alle Wortpaare und einmal ein variabler Parameter für alle Wortpaare. Alle Simulationen, die mit dem Estes Modell mit den Testkorpora durchgeführt wurden, werden im folgenden Abschnitt bezeichnet und zusammengefasst:

- *E1*: Bei dieser Simulation erhält jedes Wortpaar einen konstanten Wert für θ . θ hat den Wert 0.000021. Dieser Wert für θ wurde mit dem Kalibrierkorpus im vorherigen Kapitel ermittelt.
- *E2*: Die Ergebnisse der Simulation *E1* werden mit dem Informationsgehalt des zugehörigen Responsewortes gewichtet.

$$g_{i,j} = a_{i,j} \cdot I_j \quad (7.1)$$

- *E3*: Bei dieser Simulation erhält jedes Wortpaar einen unterschiedlichen Wert für θ . Dieser Wert wird nach folgender Formel berechnet:

$$\theta_{i,j} = 3.6E - 19 \cdot (I_i \cdot I_j)^5 \quad (7.2)$$

Das Gallistel Modell lieferte nur gute Vorhersagen, wenn die Ergebnisse nach der Berechnung noch gewichtet wurden, sonst wurden die Ergebnisse zu stark durch den Worthäufigkeitseffekt verfälscht. Die Gewichtung wurde mit dem potenzierten Informationsgehalt des Responsewortes durchgeführt (siehe Kapitel 6, Abschnitt Parameterschätzung). Für die Simulationen mit den Testkorpora wurden zwei unterschiedliche Gewichtungen verwendet. Alle Simulationen, die mit dem Gallistel Modell mit den Testkorpora durchgeführt wurden, werden im folgenden Abschnitt bezeichnet und zusammengefasst:

- *G1*: Diese Simulation verwendet eine Potenzierung von $n = 5.6$. Die gewichtete Assoziationsstärke $g_{i,j}$ eines Responsewortes j zu einem gegebenen Stimuluswort i wird nach folgender Formel berechnet:

$$g_{i,j} = a_{i,j} \cdot I_j^{5.6} \quad (7.3)$$

- *G2*: Diese Simulation verwendet eine Potenzierung von $n = 5$. Die gewichtete Assoziationsstärke $g_{i,j}$ eines Responsewortes j zu einem gegebenen Stimuluswort i wird nach folgender Formel berechnet:

$$g_{i,j} = a_{i,j} \cdot I_j^5 \quad (7.4)$$

Beim Rescorla & Wagner Modell wurden drei Simulationsvarianten verwendet. Eine Variante verwendete sowohl die Lernregel als auch die Regel für Vergessen. Desweiteren ergab sich bei der Parameterschätzung, dass ohne Anwendung der Regel für Vergessen gleich gute Ergebnisse wie mit der Regel für Vergessen erzielt werden konnten. Deswegen wurde bei der zweiten und dritten Variante auf die Regel für Vergessen verzichtet. Bei der Parameterschätzung führte eine zusätzliche Rotation des Fensterinhaltes zu besseren Ergebnissen als wenn das Fenster nicht rotiert wurde. Deswegen wurde bei der dritten Variante der Inhalt des Fensters rotiert.

Alle Simulationen, die mit dem Rescorla & Wagner Modell mit den Testkorpora durchgeführt wurden, werden im folgenden Abschnitt bezeichnet und zusammengefasst:

- *R1*: Diese Simulation verwendet die Lernregel und die Regel für Vergessen. Alle Wortpaare erhalten bei Anwendung der Lernregel einen unterschiedlichen Wert für den Parameter $\alpha_i \cdot \beta_j$:

$$\alpha_i \cdot \beta_j = 3.6E - 18 \cdot (I_i \cdot I_j)^6 \quad (7.5)$$

Bei der Anwendung der Regel für Vergessen erhalten die einzelnen Wortpaare für $\alpha_i \cdot \beta_j$ den Wert:

$$\alpha_i \cdot \beta_j = 3.6E - 20 \cdot (I_i \cdot I_j)^6 \quad (7.6)$$

- *R2*: Bei dieser Simulation wird nur die Lernregel ausgeführt. Der Parameter für Lernen wird für ein Wortpaar (i, j) nach folgender Formel berechnet:

$$\alpha_i \cdot \beta_j = 3.6E - 16 \cdot (I_i \cdot I_j)^{5.5} \quad (7.7)$$

- *R3*: Bei dieser Simulation wird ebenfalls nur die Lernregel verwendet. Zusätzlich wird aber noch der Fensterinhalt rotiert. Der Lernparameter für ein Wortpaar wird folgendermaßen berechnet:

$$\alpha_i \cdot \beta_j = 3.6E - 21 \cdot (I_i \cdot I_j)^5 \quad (7.8)$$

Insgesamt wurden 8 Simulationen mit den vier Testkorpora, 3 Simulationen mit dem Estes Modell, 2 Simulationen mit dem Gallistel Modell und 3 Simulationen mit dem Rescorla & Wagner Modell, durchgeführt. Im folgenden Abschnitt werden die Ergebnisse der 8 Simulationen, die mit den unterschiedlichen Testkorpora erzielt wurden, in tabellarischer Form dargestellt. Für jedes Textkorpus, SZ95, SZ96, SZ97 und SZ existiert eine Tabelle. Diese Tabellen enthalten in den Spalten 2 bis 5 die vier Kriterien. Die Zeilen enthalten die Bezeichnungen der unterschiedlichen Simulationen, die mit den einzelnen Modellen durchgeführt wurden. Alle Simulationen wurden mit einer Fensterlänge von 22 durchgeführt.

Die Tabelle 7.1 zeigt, dass die Simulation *E2* die besten Ergebnisse in Bezug auf das erste Kriterium, Anzahl richtig vorausgesagter Primärantworten, lieferte, wenn das Textkorpus SZ95 verwendet wurde. Die Simulation *E1* lieferte nur eine richtig vorausgesagte Primärantwort weniger als die Simulation *E2*. Beim dritten Kriterium wurde das beste Ergebnis von der Simulation *E1* erzielt. Sie lieferte aber beim zweiten Kriterium das schlechteste Ergebnis. Die Simulation *E3* konnte von den Simulationen des Estes Modells die wenigsten Primärantworten auf Rangplatz 1 richtig voraussagen, lieferte aber

SZ95	Krit. 1	Krit. 2	Krit. 3	Krit. 4
<i>E1</i>	24	270	36	52
<i>E2</i>	25	283	34	54
<i>E3</i>	21	307	33	54
<i>G1</i>	17	270	23	45
<i>G2</i>	17	277	27	47
<i>R1</i>	18	277	27	47
<i>R2</i>	17	281	26	45
<i>R3</i>	21	304	33	54

Tabelle 7.1: Die Ergebnisse der unterschiedlichen Simulationen der einzelnen Modelle in Bezug auf die vier Kriterien, wenn das Textkorpus SZ95 verwendet wurde. Mit *E2* konnten die meisten Primärantworten richtig vorausgesagt werden. Die Simulationen *E3* und *R3* lieferten fast die gleichen Ergebnisse in Bezug auf die vier Kriterien. Sie lieferten die besten Ergebnisse in Bezug auf das zweite Kriterium. Beim dritten und vierten Kriterium schnitten die Simulationen des Estes Modells und die Simulation *R3* des Rescorla & Wagner Modells am besten ab.

das beste Ergebnis in Bezug auf das zweite Kriterium. Die beiden Simulationen des Gallistel Modells und die beiden Simulationen des Rescorla & Wagner Modells *R1* und *R2*, bei denen nicht der Fensterinhalt rotiert wurde, lieferten die schlechtesten Ergebnisse in Bezug auf die Kriterien 1, 3 und 4. Die Simulation *R3*, bei der eine Rotation des Fensterinhaltes erfolgte, produzierte die besten Ergebnisse in Bezug auf die vier Kriterien von den drei Simulationen des Rescorla & Wagner Modells. Sie waren genauso gut wie die Ergebnisse von *E3*. Die beiden Simulationen *E3* und *R3* lieferten die beste Ergebnisse in Bezug auf das zweite Kriterium, das alle Responsewörter der Teilnehmer berücksichtigte.

In der Tabelle 7.2 werden die Ergebnisse der Simulationen dargestellt, wenn das Textkorpus SZ96 als Lernmaterial verwendet wurde.

In Bezug auf das erste Kriterium, Anzahl richtig vorausgesagter Primärantworten, konnte wieder eine Simulation des Estes Modells die besten Voraussagen machen. Diesmal war es die Simulation *E1*. Die Simulation *E2* des Estes Modells lieferte das zweitbeste Ergebnis in Bezug auf das erste Kriterium, aber das beste Ergebnis in Bezug auf das dritte Kriterium. Während in Bezug auf das zweite Kriterium und vierte Kriterium die Simulation *E3* und die Simulation des Rescorla & Wagner Modells *R3* die besten Ergebnisse lieferten. Die beiden Simulationen des Gallistel Modells produzierten die schlechtesten Ergebnisse in Bezug auf alle vier Kriterien. Etwas

SZ96	Krit. 1	Krit. 2	Krit. 3	Krit. 4
<i>E1</i>	25	260	35	51
<i>E2</i>	23	274	37	53
<i>E3</i>	19	302	36	55
<i>G1</i>	17	263	28	45
<i>G2</i>	18	265	27	46
<i>R1</i>	19	277	30	47
<i>R2</i>	17	277	29	48
<i>R3</i>	19	303	36	55

Tabelle 7.2: Die Ergebnisse der unterschiedlichen Simulationen der einzelnen Modelle in Bezug auf die vier Kriterien, wenn das Textkorpus SZ96 verwendet wurde. *E1* konnte zwar die meisten Primärantworten voraussagen, lieferte aber das schlechteste Ergebnis in Bezug auf das zweite Kriterium. *E3* und *R3* produzierten fast gleiche Ergebnisse in Bezug auf die vier Kriterien. Die beiden Simulationen des Gallistel Modells unterschieden sich in den erzielten Ergebnissen nur wenig voneinander. Die besten Ergebnisse in Bezug auf das zweite Kriterium wurden mit den Simulationen *E3* und *R3* erzielt. Sie produzierten auch das beste Ergebnis in Bezug auf das vierte Kriterium.

besser schnitten die Simulationen des Rescorla & Wagner Modells *R1* und *R2* ab. *R3* produzierte die besseren Ergebnisse von den drei Simulationen des Rescorla & Wagner Modells in Bezug auf das zweite, dritte und vierte Kriterium.

Die Ergebnisse der Simulationen mit dem Textkorpus SZ97 werden in der Tabelle 7.3 präsentiert.

Die Simulation *E3* lieferte das beste Ergebnis in Bezug auf das erste Kriterium. Beim zweiten Kriterium produzierten wieder *E3* und *R3* die besten Voraussagen. Beim dritten und vierten Kriterium erzielten *E1* und *E2* die besten Ergebnisse. Insgesamt unterschieden sich die erzielten Ergebnisse aller drei Modelle im Bezug auf das erste Kriterium weniger voneinander als bei den vorherigen beiden Textkorpora. Die Simulationen des Gallistel Modells *G1* und *G2* und die Simulationen des Rescorla & Wagner Modells *R1* und *R2*, die bisher immer schlechter beim ersten Kriterium abgeschnitten hatten, produzierten jetzt ähnlich gute Ergebnisse wie die anderen Simulationen. Alle drei Modelle konnten jetzt mindestens in einer Simulation 23 Primärantworten richtig voraussagen. Die Teilnehmer produzierten in der Erhebung von Russell & Meseck im Mittel fast gleich viele Primärantworten. Dies waren 22,5 richtige Primärantworten.

Fast alle Simulationen konnten die Ergebnisse in Bezug auf das zweite

SZ97	Krit. 1	Krit. 2	Krit. 3	Krit. 4
<i>E1</i>	23	269	35	51
<i>E2</i>	22	286	35	52
<i>E3</i>	24	312	33	50
<i>G1</i>	23	285	30	46
<i>G2</i>	20	285	30	45
<i>R1</i>	23	294	30	45
<i>R2</i>	22	297	29	45
<i>R3</i>	23	313	33	50

Tabelle 7.3: Die Ergebnisse der unterschiedlichen Simulationen der einzelnen Modelle in Bezug auf die vier Kriterien, wenn das Textkorpus SZ97 verwendet wurde. Mit der Simulation *E3* konnte die meisten Primärantworten vorausgesagt werden. Die besten Ergebnisse in Bezug auf das zweite Kriterium wurden mit den Simulationen *E3* und *R3* erzielt. Alle Simulationen konnten mindestens 20 Primärantworten auf Rangplatz 1 richtig voraussagen. Mindestens eine Simulation von jedem Modell konnte 23 Primärantworten auf Rangplatz 1 richtig voraussagen.

Kriterium verbessern, wenn das Textkorpus SZ97 verwendet wurde.

In der Tabelle 7.4 werden die Ergebnisse aufgeführt, wenn alle vier Jahrgänge SZ zusammen als Lernmaterial für die Simulationen verwendet wurden.

Die meisten Primärantworten wurden mit der Simulation *E2* des Estes Modells richtig vorausgesagt. Beim zweiten Kriterium produzierten die Simulationen *R2*, *R3* und *E3* die besten Ergebnisse. Wie bei den vorherigen Simulationen produzierten die Simulation *E3* des Estes Modells und die Simulation *R3* des Rescorla & Wagner Modells fast die gleichen Ergebnisse in Bezug auf die vier Kriterien. Bei dem Kriterium 2 schnitten die Simulationen *E1* und *E2* am schlechtesten ab. Dafür lieferte die Simulation *E2* die besten Ergebnisse in Bezug auf das dritte Kriterium. Beim vierten Kriterium wurde das beste Ergebnis von *G2* erzielt. Die Voraussagen aller Modelle verbesserten sich gegenüber den Voraussagen, die bisher mit den Zeitungskorpora erzielt wurden. Dies war auch zu erwarten, da der letzte Teil des Textkorpus SZ den Kalibrierkorpus SZ98 enthielt.

Beim zweiten Kriterium konnte bei allen Modellen eine deutliche Verbesserung beobachtet werden. Bei keinem Textkorpus wurden so gute Ergebnisse in Bezug auf das zweite Kriterium erzielt. Diese guten Ergebnisse in Bezug auf das zweite Kriterium wurden nicht einmal mit dem Kalibrierkorpus erreicht. Bei allen Modellen waren mindestens 300 Nennungen der Teilnehmer auf den

SZ	Krit. 1	Krit. 2	Krit. 3	Krit. 4
<i>E1</i>	26	307	37	56
<i>E2</i>	29	312	39	56
<i>E3</i>	25	347	36	55
<i>G1</i>	23	340	33	53
<i>G2</i>	24	341	32	57
<i>R1</i>	24	333	35	56
<i>R2</i>	24	345	32	54
<i>R3</i>	25	346	36	56

Tabelle 7.4: Die Ergebnisse der unterschiedlichen Simulationen der einzelnen Modelle in Bezug auf die vier Kriterien, wenn alle vier Jahrgänge SZ zusammen als Textkorpus verwendet wurden. *E2* konnte die meisten Primärantworten richtig voraussagen, lieferte aber das zweit schlechteste Ergebnis in Bezug auf das zweite Kriterium. Dafür erzielte *E2* aber beim dritten Kriterium wieder das beste Ergebnis. Die besten Ergebnisse in Bezug auf das zweite Kriterium wurden von *E3* und *R3* erzielt. Das beste Ergebnis in Bezug auf das vierte Kriterium wurde von der Simulation *G2* erzielt.

ersten zehn Rangplätzen. Berechnet man aus diesem Wert die durchschnittliche Anzahl der Nennungen der Teilnehmer (Mittelwert) pro Stimuluswort auf den ersten zehn Rangplätzen, so waren mindestens drei Nennungen pro Stimuluswort auf den ersten zehn Rangplätzen zu finden. Die Primärantwort wurde hierbei als Nennung nicht berücksichtigt.

Bei allen Modellen war für über die Hälfte aller Stimuluswörter die zugehörige Primärantwort auf den ersten fünf Rangplätzen zu finden, wenn das Textkorpus SZ als Lernmaterial verwendet wurde.

7.1 Kriterium 1

In der Abbildung 7.1 sind die Anzahl richtig vorausgesagter Primärantworten der einzelnen Simulationen in Abhängigkeit von dem verwendeten Textkorpus aufgeführt. Für jede der 8 Simulationen existieren vier Balken in der Abbildung 7.1. Die Farbe der Balken gibt den verwendeten Textkorpus an. Das Textkorpus SZ95 wird durch weiße Balken repräsentiert, das Textkorpus SZ96 durch hellgraue Balken, das Textkorpus SZ97 durch dunkelgraue Balken und das Textkorpus SZ durch schwarze Balken. Die Zahl, die in der Abbildung 7.1 über den Balken steht, gibt die Anzahl richtig vorausgesagter Primärantworten, die mit zugehörigen Textkorpus erzielt wurden, an.

Bei allen Simulationen wurde das beste Ergebnis in Bezug auf das erste Kriterium erzielt, wenn das Textkorpus SZ verwendet wurde (siehe schwarze Balken). Die Simulationen *E3* und *R3* produzierten bei allen vier Textkorpora fast dieselbe Anzahl richtig vorausgesagter Primärantworten. Mit der Simulation *E2* des Estes Modells wurden die meisten Primärantworten vorausgesagt, wenn die gesamte SZ als Lernmaterial verwendet wurde. Die beiden Simulationen *E1* und *E2* produzierten die besten Ergebnisse in Bezug auf das erste Kriterium, wenn die beiden kleineren Textkorpora SZ95 und SZ96 als Lernmaterial verwendet wurden.

Wurden die Textkorpora SZ95 und SZ96 für die Berechnung der Wortassoziationen verwendet, so variierte die Anzahl richtig vorausgesagter Primärantworten am stärksten bei den unterschiedlichen Simulationen, sie variierte zwischen 17 und 25 richtig vorausgesagter Primärantworten (siehe weiße und hellgraue Balken). Beim Textkorpus SZ97 variierte die Anzahl richtig vorausgesagter Primärantworten am wenigsten, sie variierte zwischen 20 und 24 richtig vorausgesagter Primärantworten (siehe dunkelgraue Balken). Dies war ungefähr dieselbe Anzahl von Primärantworten, die im Mittel ein Teilnehmer in der Erhebung von Russell & Meseck produziert hatte. Er produzierte 22,5 richtige Primärantworten.

Die Anzahl richtig vorausgesagter Primärantworten variierte zwischen 23 und 29 bei den unterschiedlichen Simulationen, wenn die gesamte SZ als Lernmaterial verwendet wurde (siehe schwarze Balken).

Bei den ersten beiden Textkorpora SZ95 und SZ96 lieferten die Simulationen des Gallistel Modells und die Simulationen des Rescorla & Wagner Modells *R1* und *R2* die schlechtesten Voraussagen in Bezug auf das erste Kriterium. Sie konnten nicht mindestens 20 Primärantworten auf den ersten Rangplatz richtig voraussagen.

Die besten Ergebnisse in Bezug auf das erste Kriterium wurden bei allen verwendeten Textkorpora immer von einer Simulation des Estes Modells erzielt, aber es war nicht immer dieselbe Simulation. Bei der SZ95 war dies *E2*, bei der SZ96 *E1*, bei der SZ97 *E3* und bei der gesamten SZ war es die Simulation *E2* (siehe Abbildung 7.1).

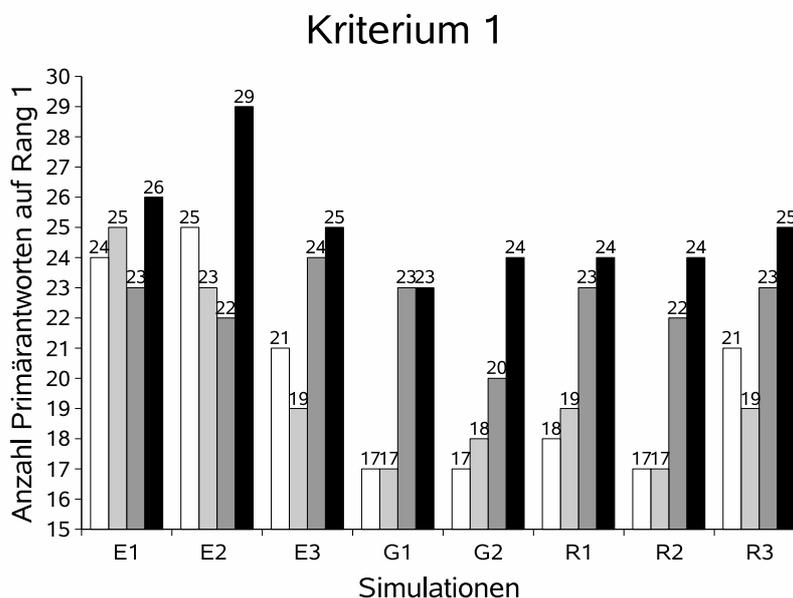


Abbildung 7.1: Für jede Simulation wird die Anzahl richtig vorausgesagter Primärantworten in Abhängigkeit von dem verwendeten Lernmaterial dargestellt.

Das verwendete Lernmaterial wird durch die unterschiedlich farbigen Balken repräsentiert. Bei der Anzahl richtig vorausgesagter Primärantworten über den weißen Balken wurde das Textkorpus SZ95 verwendet, entsprechend wurde bei den hellgrauen Balken das Textkorpus SZ96, bei den grauen Balken das Textkorpus SZ97 und bei den schwarzen Balken das Textkorpus SZ verwendet. Bei allen Simulationen wurde das beste Ergebnis in Bezug auf das erste Kriterium erzielt, wenn das Textkorpus SZ verwendet wurde (siehe schwarze Balken).

Die Simulationen *E3* und *R3* produzierten bei allen vier Textkorpora fast dieselbe Anzahl richtig vorausgesagter Primärantworten. Mit der Simulation *E2* des Estes Modells wurden die meisten Primärantworten vorausgesagt, wenn die gesamte SZ als Lernmaterial verwendet wurde.

Im Vergleich zu den Simulationsergebnissen produzierte ein Teilnehmer in der Erhebung von Russell & Meseck 22,5 richtige Primärantworten. Diese Anzahl Primärantworten konnten die Simulationen *E1* und *E2* bei allen verwendeten Textkorpora auch voraussagen. Die anderen Simulationen konnten diese Anzahl Primärantworten voraussagen, wenn das Textkorpus SZ97 oder das Textkorpus SZ verwendet wurde (G2 nur wenn das Textkorpus SZ verwendet wurde).

7.2 Kriterium 2

In der Abbildung 7.2 werden die Ergebnisse, die mit den unterschiedlichen Simulationen der einzelnen Modelle in Abhängigkeit von dem verwendeten Textkorpus vorhergesagt wurden, in Bezug auf das zweite Kriterium dargestellt. Bei allen Simulationen wurde das beste Ergebnis in Bezug auf das zweite Kriterium erzielt, wenn das Textkorpus SZ verwendet wurde (siehe schwarze Balken).

Wurde das Textkorpus SZ97 als Lernmaterial verwendet, so konnten fast alle Simulationen ihr zweitbestes Ergebnis in Bezug auf das zweite Kriterium erzielen (siehe dunkelgraue Balken).

Fast alle Simulationen erzielten das schlechteste Ergebnis in Bezug auf das zweite Kriterium, wenn das Textkorpus SZ96 verwendet wurde (siehe hellgraue Balken).

Die Simulationen *E3*, *R2* und *R3* lieferten die besten Ergebnisse in Bezug auf das zweite Kriterium, wenn die gesamte SZ als Lernmaterial verwendet wurde.

Bei allen verwendeten Textkorpora schnitt die Simulation *E1* beim Kriterium 2 am schlechtesten ab.

Bei dem Kriterium 2 schnitt die Simulation *E3* wesentlich besser ab als die anderen beiden Simulationen des Estes Modells (siehe in der Abbildung 7.2 die vier Balken über *E3*). Bei Verwendung des Textkorpus SZ95 und des Textkorpus SZ lieferte die Simulation *E3* sogar das beste Ergebnis in Bezug auf das zweite Kriterium, bei den anderen Textkorpora das zweitbeste Ergebnis.

Die Simulationen *E3* und *R3* unterschieden sich beim zweiten Kriterium nur sehr wenig voneinander. Sie konnten bei allen verwendeten Textkorpora mindestens 300 Nennungen der Teilnehmer auf den ersten zehn Rangplätzen richtig voraussagen. Im Mittel wurden von diesen beiden Simulationen drei Responsewörter der Teilnehmer pro Stimuluswort auf den ersten zehn Rangplätzen richtig vorausgesagt.

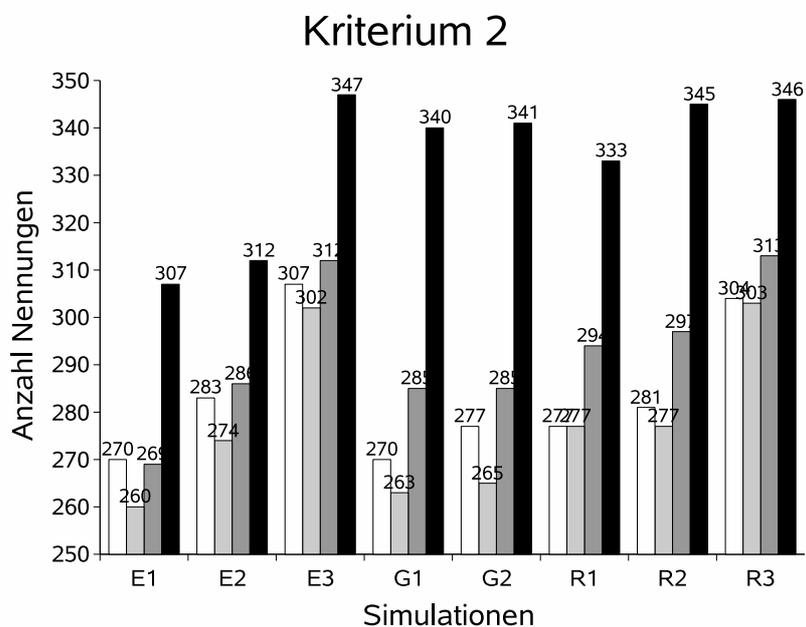


Abbildung 7.2: Für jede Simulation wird die Anzahl der Nennungen der Teilnehmer im Assoziationsversuch auf den ersten zehn Rangplätzen in Abhängigkeit von dem verwendeten Lernmaterial dargestellt. Die Primärantwort wird dabei nicht als Nennungen betrachtet.

Das verwendete Lernmaterial wird durch die unterschiedlich farbigen Balken repräsentiert. Die weißen Balken gehören zu den Ergebnissen, die mit dem Textkorpus SZ95 erzielt wurden, entsprechend wurde bei den hellgrauen Balken das Textkorpus SZ96, bei den grauen Balken das Textkorpus SZ97 und bei den schwarzen Balken das Textkorpus SZ verwendet. Bei allen Simulationen wurde das beste Ergebnis in Bezug auf das zweite Kriterium erzielt, wenn das Textkorpus SZ verwendet wurde (siehe schwarze Balken).

Die Simulationen *E3*, *R2* und *R3* lieferten die besten Ergebnisse in Bezug auf das zweite Kriterium, wenn die gesamte SZ als Lernmaterial verwendet wurde. Sie konnten mindestens 345 Nennungen der Teilnehmer auf den ersten zehn Rangplätzen richtig voraussagen. Im Mittel wurden von diesen drei Simulationen 3,4 Responsewörter der Teilnehmer pro Stimuluswort auf den ersten zehn Rangplätzen richtig vorausgesagt, wenn das Textkorpus SZ verwendet wurde.

7.3 Kriterium 3

In der Abbildung 7.3 wird die Anzahl richtig vorausgesagter Primärantworten, die mit den unterschiedlichen Simulationen der einzelnen Modelle in Abhängigkeit von dem verwendeten Textkorpus auf Rang 1 oder Rang 2 vorhergesagt wurden, dargestellt.

Die Simulationen des Estes Modells lieferten in Bezug auf das dritte Kriterium bei allen verwendeten Textkorpora bessere oder gleich gute Ergebnisse wie die anderen Modelle. Beim dritten Kriterium wurden von den Simulationen des Estes Modells mindestens 33 Primärantworten auf dem ersten oder zweiten Rangplatz richtig vorausgesagt.

Bei allen Simulationen wurde das beste Ergebnis in Bezug auf das dritte Kriterium erzielt, wenn das Textkorpus SZ verwendet wurde (siehe schwarze Balken). Alle Simulationen konnten mindestens 32 Primärantworten auf den ersten beiden Rangplätzen voraussagen, wenn das Textkorpus SZ als Lernmaterial verwendet wurde.

Wurde das Textkorpus SZ97 als Lernmaterial verwendet, so konnten von allen Simulationen mindestens 29 Primärantworten auf den ersten beiden Rangplätzen richtig vorausgesagt werden.

Das beste Ergebnis in Bezug auf das dritte Kriterium wurde mit der Simulation *E2* erzielt, wenn das Textkorpus SZ verwendet wurde.

Die Simulationen *E3* und *R3* erzielten dasselbe Ergebnis in Bezug auf das dritte Kriterium bei allen verwendeten Textkorpora.

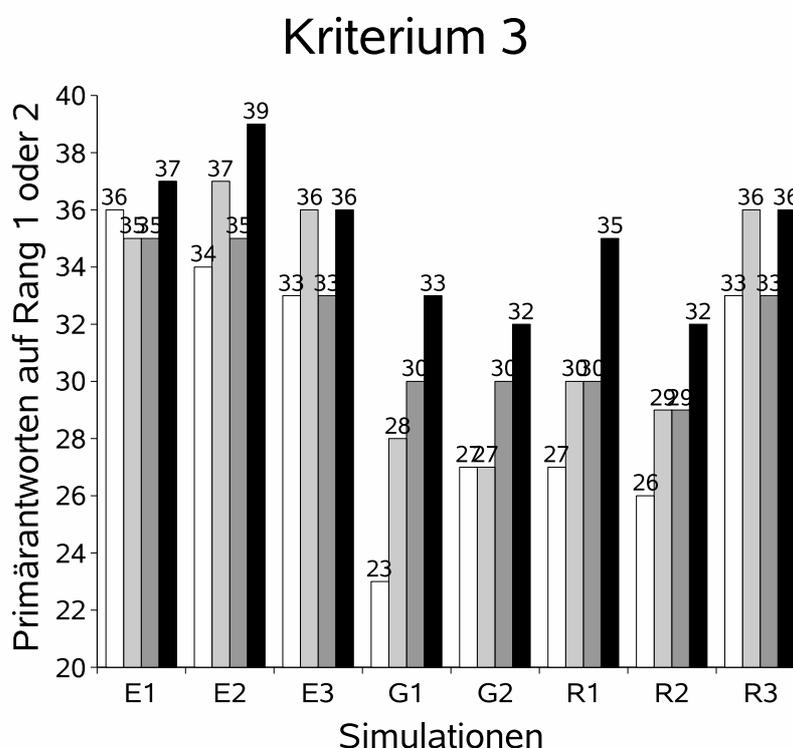


Abbildung 7.3: Für jede Simulation wird die Anzahl richtig vorausgesagter Primärantworten auf den ersten beiden Rangplätzen in Abhängigkeit von dem verwendeten Lernmaterial dargestellt.

Das verwendete Lernmaterial wird durch die unterschiedlich farbigen Balken repräsentiert. Die weißen Balken gehören zu den Ergebnissen, die mit dem Textkorporus SZ95 erzielt wurden, entsprechend wurde bei den hellgrauen Balken das Textkorporus SZ96, bei den grauen Balken das Textkorporus SZ97 und bei den schwarzen Balken das Textkorporus SZ verwendet. Bei allen Simulationen wurde das beste Ergebnis in Bezug auf das dritte Kriterium erzielt, wenn die gesamte SZ als Textkorporus verwendet wurde (siehe schwarze Balken). Die Simulation *E2* erzielte das beste Ergebnis in Bezug auf das dritte Kriterium, wenn das Textkorporus SZ und das Textkorporus SZ96 verwendet wurde (siehe hellgraue und schwarze Balken). Die Simulation *E1* erzielte das beste Ergebnis in Bezug auf das dritte Kriterium, wenn das Textkorporus SZ95 verwendet wurde (siehe weiße Balken). Wurde das Textkorporus SZ97 als Lernmaterial verwendet, so erzielten die Simulationen *E1* und *E2* das beste Ergebnis in Bezug auf das dritte Kriterium (siehe dunkelgraue Balken).

7.4 Kriterium 4

In der Abbildung 7.4 wird die Anzahl richtig vorausgesagter Primärantworten, die mit den unterschiedlichen Simulationen der einzelnen Modelle in Abhängigkeit von dem verwendeten Textkorpus auf Rang 1 bis Rang 5 vorhergesagt wurden, dargestellt. Die Simulationen des Estes Modells und die Simulation *R3* konnten bei allen verwendeten Textkorpora mindestens für die Hälfte aller Stimuluswörter die zugehörige Primärantwort auf den ersten fünf Rangplätzen richtig voraussagen. Die anderen Simulationen konnten dies nur wenn das Textkorpus SZ als Lernmaterial verwendet wurde.

Alle Simulationen konnten bei den verwendeten Textkorpora mindestens 45 Primärantworten auf den ersten fünf Rangplätzen richtig voraussagen.

Bei dem vierten Kriterium lieferten die Simulationen des Estes Modells bei allen verwendeten Textkorpora bessere oder fast gleich gute Ergebnisse wie die anderen Modelle.

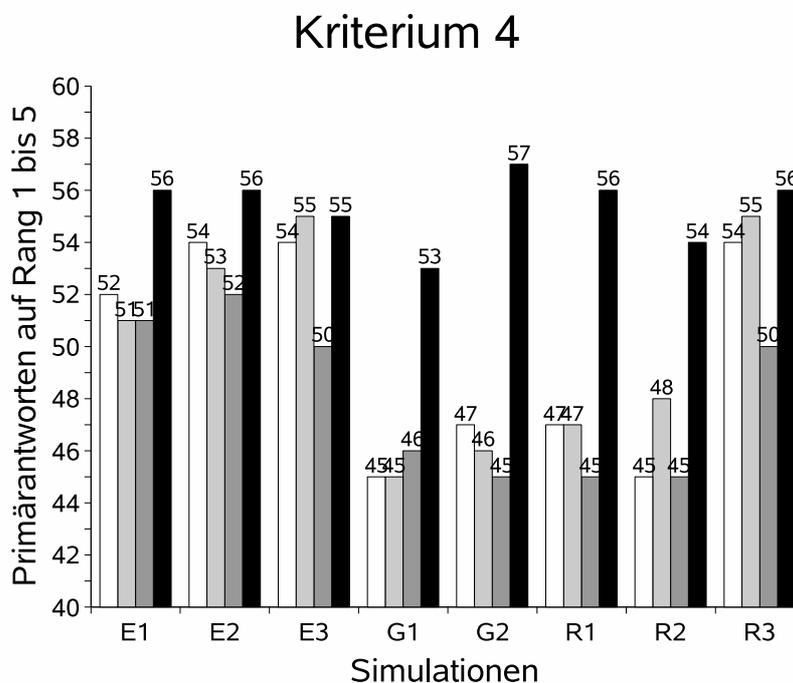


Abbildung 7.4: Für jede Simulation wird die Anzahl richtig vorausgesagter Primärantworten auf den ersten fünf Rangplätzen in Abhängigkeit von dem verwendeten Lernmaterial dargestellt.

Das verwendete Lernmaterial wird durch die unterschiedlich farbigen Balken repräsentiert. Die weißen Balken gehören zu den Ergebnissen, die mit dem Textkorpus SZ95 erzielt wurden, entsprechend wurde bei den hellgrauen Balken das Textkorpus SZ96, bei den grauen Balken das Textkorpus SZ97 und bei den schwarzen Balken das Textkorpus SZ verwendet. Bei allen Simulationen wurde das beste Ergebnis in Bezug auf das dritte Kriterium erzielt, wenn die gesamte SZ als Textkorpus verwendet wurde (siehe schwarze Balken).

Die drei Simulationen des Estes Modells und die Simulation *R3* des Rescorla & Wagner Modells konnten bei allen verwendeten Textkorpora für die Hälfte der Stimuluswörter die Primärantwort auf den ersten fünf Rangplätzen richtig voraussagen (auf den Balken über *E1*, *E2*, *E3* und *R3* steht mindestens ein Wert von 50). Die anderen Simulationen konnten dies nur wenn das Textkorpus SZ verwendet wurde. Das beste Ergebnis in Bezug auf das vierte Kriterium wurde mit der Simulation *G2* erzielt, wenn das Textkorpus SZ als Lernmaterial verwendet wurde.

7.5 Vergleich und Interpretation der Simulationsergebnisse

Beim Rescorla & Wagner Modell schnitt die Simulation *R3* bei den ersten beiden Korpora SZ95 und SZ96 immer besser ab als die anderen beiden Simulationen. Die Simulation *R3* des Rescorla & Wagner Modells unterschied sich von den anderen beiden Simulationen dadurch, dass der Fensterinhalt rotiert wurde. Dies hatte ungefähr denselben Effekt als wenn das Textkorpus vergrößert wurde. Da die beiden Korpora SZ95 und SZ96 die kleinsten Textkorpora sind, würden die besseren Ergebnisse von *R3* in Bezug auf die vier Kriterien dafür sprechen, dass eine gewisse Korporagröße für gute Ergebnisse bei den Simulationen des Rescorla & Wagner Modells benötigt wurde. Denn wenn das Textkorpus größer wurde, dann wurden auch die Ergebnisse von *R1* und *R2* besser. Die erzielten Ergebnisse von *R3* und den anderen beiden Simulationen *R1* und *R2* unterschieden sich bei den größeren Textkorpora nicht mehr so stark voneinander.

Die Simulationen *E1* und *E2* lieferten auch bei den beiden kleineren Textkorpora gute Ergebnisse in Bezug auf die drei Kriterien, die die Anzahl richtig vorausgesagter Primärantworten berücksichtigten. Die Simulationen der anderen Modelle verbesserten sich, wenn das Textkorpus größer wurde. Dies kann einmal daran gelegen haben, dass mit dem größten Textkorpus SZ98 von den drei Textkorpora SZ95, SZ96 und SZ97 die freien Parameter ermittelt wurden, oder es kann daran gelegen haben, dass eine gewisse Korporagröße für die Simulationen benötigt wurde.

Vor allem für das zweite Kriterium scheint eine gewisse Größe des Textkorpus benötigt zu werden, denn mit dem größten Textkorpus wurden die besten Ergebnisse in Bezug auf dieses Kriterium erzielt. Diese guten Ergebnisse in Bezug auf das zweite Kriterium, die mit dem größten Textkorpus SZ von allen Simulationen erzielt wurden, konnten nicht einmal mit dem Kalibrierkorpus erzielt werden.

Es zeigte sich, dass die Simulationen *E3* und *R3*, bei der der Fensterinhalt rotiert wurde, fast identische Ergebnisse in Bezug auf die vier Kriterien bei allen verwendeten Textkorpora lieferten. Dies war auch schon beim Kalibrierkorpus festgestellt worden.

Mit allen drei Lernmodellen konnte das Erlernen von Wortassoziationen simuliert werden. Insgesamt schnitten bei der Voraussage von Wortassoziationen die Simulationen des Estes Modells bei allen vier Textkorpora etwas besser als die Simulationen der anderen Modelle ab. Da das Estes Modell etwas besser als die anderen Modelle bei der Voraussage von deutschen Wortassoziationen abgeschnitten hatte, wurde es im nächsten Kapitel zur Voraus-

sage von englischen Wortassoziationen verwendet.

7.6 Fazit

Am Ende von Kapitel 3 war die Vermutung aufgestellt worden, dass Wortassoziation die Ergebnisse von Lernprozessen sind. Die obigen Ergebnisse bestätigten diese Vermutung. Das Erlernen von Wortassoziationen konnte mit den drei Lernmodellen simuliert werden. Für das Erlernen von Wortassoziationen reichte ein einfaches Lernmodell (Estes Modell), das Lernen und Interferenz (Vergessen) erklären kann, aus. Komplexere Lernmodelle, wie das Rescorla & Wagner Modell und das Gallistel Modell, die nicht nur Lernen und Interferenz erklären können, sondern auch komplexere Phänomene wie Blocking, Überlagerung und Hemmung, die in klassischen Konditionierungsexperimenten beobachtet wurden, produzierten beim Erlernen von Wortassoziationen keine besseren Ergebnisse in Bezug auf die vier Kriterien.

Kapitel 8

Englische Wortassoziationen

In diesem Abschnitt wird die Vorhersage von englischen Wortassoziationen beschrieben. Als Assoziationsnormen dienten, die im Kapitel 2 beschriebenen englischen Assoziationsnormen von Nelson, McEvoy und Schreiber (Nelson, McEvoy & Schreiber, 1999). Als Textkorpus wurde das British National Korpus (BNC), das schon Wettler und Rapp (Wettler, Rapp & Sedlmeier, 2005) für den Aufbau ihres Wortnetzes verwendet hatten, verwendet. Dieses Textkorpus ist etwas kleiner als die vier Jahrgänge SZ zusammen. Das Textkorpus BNC besteht aus ca. 100 Millionen Wörter, während die vier Jahrgänge SZ ca. 123 Millionen Wörter enthalten. Das Vokabular bestand aus allen Stimuluswörtern und den zugehörigen Responsewörtern aus den Assoziationsnormen von Nelson, McEvoy und Schreiber (Nelson, McEvoy & Schreiber, 1999), die im BNC mindestens zehn mal vorkamen. Insgesamt waren dies 9126 Wörter. In den Assoziationsnormen sind auch sogenannte Stop Wörter (z.B. *the, a, and, of, or* usw.) enthalten, die Griffiths und Steyvers bei ihrer Vorhersage von Wortassoziationen nicht verwendet hatten. Ihr Vokabular enthielt deswegen nur 4543 Wörter aus den Assoziationsnormen.

Von den 5019 Stimuluswörtern, von denen die Assoziationsnormen erhoben wurden, kamen 4958 Stimuluswörter wenigstens zehn mal im BNC vor. Somit enthielt das Vokabular 4958 Stimuluswörter. Die Berechnung der Assoziationen zu einer so große Anzahl von Stimuluswörtern verursachte einen großen Rechenaufwand. Deswegen wurde die Vorhersage von englischen Wortassoziationen nur mit einem Lernmodell durchgeführt. Dies war das Estes Modell, da es im vorherigen Abschnitt leicht bessere Ergebnisse bei der Voraussage der deutschen Wortassoziationen in Bezug auf die vier Kriterien geliefert hatte.

Das Textkorpus BNC wurde nicht für die Simulationen lemmatisiert, da die englischsprachigen Assoziationsnormen auch flektierte Reizwörter enthalten. Dies ist bei den deutschen Assoziationsnormen nicht der Fall. Außerdem

enthält die englische Sprache nicht so viele flektierte Formen wie die deutsche Sprache, so dass sie sich für den Versuch eignet, Assoziationsberechnungen ohne vorherige Lemmatisierung durchzuführen.

Beim Einlesen des Textkorpus wurden alle Buchstaben in Großbuchstaben umgewandelt, da die Assoziationsnormen auch in Großbuchstaben vorlagen.

Die freien Parameter, die in dem Kapitel 6 für das Estes Modell ermittelt wurden, konnten nicht einfach übernommen werden. Dies lag einmal daran, dass vor der Durchführung der Simulationen keine Lemmatisierung erfolgte, die zu einer Erhöhung der Worthäufigkeiten geführt hätte. Ein weiterer Grund dafür, dass die Parameter aus Kapitel 6 nicht übernommen werden konnten, könnte an dem größeren Wortschatz der englischen Sprache gelegen haben. Deshalb wurde nochmal eine Parameterschätzung für das Estes Modell durchgeführt, wobei als Kalibrierkorpus das Korpus BNC verwendet wurde.

Zur Parameterschätzung und Simulation des Estes Modells wurden drei unterschiedliche Umsetzungen des Estes Modells verwendet. Dies sind die im Kapitel 7 beschriebenen Simulationen *E1*, *E2* und *E3*. Bei der Simulation *E1* wird ein konstantes θ für alle Wortpaare verwendet. Der Wert für θ musste neu bestimmt werden. Für die Simulation *E2* mussten nur die Ergebnisse von *E1* mit dem Informationsgehalt des jeweiligen Responsewortes gewichtet werden. Für *E3* musste der Faktor und das n aus Gleichung 8.1 neu bestimmt werden.

Zunächst wurde mit dem Hill- Climbing Algorithmus, der im Kapitel 6 beschrieben wurde, ein konstanter Lernparameter für alle Wortpaare ermittelt, der die besten Ergebnisse in Bezug auf die vier Kriterien lieferte. Aus Rechenzeitgründen wurde die Anzahl der durchgeführten Simulationen pro Simulationsreihe auf 7 beschränkt. Diese Einschränkung führte dazu, dass eine Simulationsreihe nur bis zu einer Fensterlänge von 50 durchgeführt wurde und, dass die Fensterlänge für die nächste Simulation jeweils um 5 statt um 2 erhöht wurde. Alle Simulationsreihen begannen erst mit einer Fensterlänge von 20 statt mit einer Fensterlänge von 12. Eine Simulationsreihe bestand somit aus den Simulationen mit den Fensterlängen 20, 25, 30, 35, 40, 45 und 50.

Die Rechenzeit einer Simulation mit einer Fensterlänge von 20 benötigte eine Rechenzeit von 7 Stunden und 30 Minuten.

In der Tabelle 8.1 sind die Ergebnisse, die mit einem konstanten θ für alle Wortpaare erzielt wurden, aufgeführt. Das beste Ergebnis in Bezug auf alle vier Kriterien wurde bei einem θ von $2E - 6$ erzielt. Bei diesem Wert wurden zwar 2 Primärantworten weniger vorausgesagt als bei einem θ von $1.9E - 6$, bei dem die maximale Anzahl richtig vorausgesagter Primärantworten er-

θ	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$2E - 4$	289	4460	446	749	20
$2E - 5$	450	6448	682	1094	30
$1E - 6$	534	6207	816	1236	50
$1.5E - 6$	549	6570	838	1267	45
$1.6E - 6$	547	6640	829	1282	40
$1.7E - 6$	546	6674	826	1282	40
$1.8E - 6$	546	6692	831	1291	40
$1.9E - 6$	550	6707	833	1292	40
$2E - 6$	548	6752	831	1297	40
$2.5E - 6$	533	6815	833	1311	40
$3E - 6$	536	6836	831	1302	40
$5E - 6$	504	6843	799	1275	40
$7E - 6$	485	6866	779	1243	35
$9E - 6$	479	6865	755	1222	30

Tabelle 8.1: Simulation von englischen Wortassoziationen mit dem Estes Modell: Es werden die Ergebnisse in Bezug auf die vier Kriterien in Abhängigkeit von dem Lernparameter θ , der für alle Wortpaare konstant gehalten wurde, aufgeführt.

Die besten Ergebnisse in Bezug auf das erste Kriterium, maximale Anzahl richtig vorausgesagter Primärantworten auf Rangplatz 1, wurden im Wertebereich von $1.5E - 6$ bis $2E - 6$ für θ erzielt. Für das zweite Kriterium wurden die besten Ergebnisse im obigen Wertebereich bei einem Wert von $2E - 6$ erzielt. Bessere Ergebnisse in Bezug auf das zweite Kriterium wurden im Wertebereich von $3E - 6$ bis $9E - 6$ erzielt, dabei sank aber die maximale Anzahl richtig vorausgesagter Primärantworten. Bei einem Maximalwert von 550 wurden 11 Prozent der Primärassoziationen richtig vorhergesagt.

reicht wurde, aber bei dem zweiten Kriterium wurden dafür ca. 50 Antworten der Teilnehmer mehr vorausgesagt. Dieses Ergebnis wurde mit einer Fensterlänge von 40 erzielt.

Bei den Simulationen mit einem konstanten θ war ein deutlicher Worthäufigkeitseffekt zu beobachten. Dies war auch schon bei der Simulation von deutschen Wortassoziationen mit einem konstanten θ beobachtet worden. Zur Verringerung des Worthäufigkeitseffekts wurden bei der Simulation $E2$ die Ergebnisse, die mit einem konstanten θ erzielt wurden, anschließend noch mit dem Informationsgehalt des zugehörigen Responsewortes gewichtet. Für die Gewichtung wurden die Ergebnisse, die bei den vorherigen Simulationen mit dem Parameter $2E - 6$ erzielt wurden, verwendet. Die besten Ergebnisse

in Bezug auf die vier Kriterien wurden bei der Gewichtung aber erst bei einer Fensterlänge von 50 erzielt, und zwar waren dies 544 richtig vorhergesagte Primärantworten, 7085 Nennungen aller Teilnehmer ohne die Primärantwort auf den Rangplätzen 1 bis 10, 863 richtig vorhergesagte Primärantworten auf Rangplatz 1 oder 2 und 1342 richtig vorhergesagte Primärantworten auf den ersten fünf Rangplätzen. Durch die Gewichtung verschlechterten sich die Ergebnisse in Bezug auf das erste Kriterium leicht (von 548 auf 544). Alle anderen Kriterien (Kriterium 2 bis 4) erzielten jedoch durch die Gewichtung bessere Ergebnisse. Das Ergebnis in Bezug auf das Kriterium 2 konnte sich von 6752 vorhergesagter Nennungen auf 7085 vorhergesagter Nennungen der Teilnehmer verbessern. Die Anzahl richtig vorausgesagter Primärantworten auf Rangplatz 1 oder 2 erhöhte sich durch die Gewichtung von 831 auf 863. Die Anzahl richtig vorausgesagter Primärantworten auf den ersten fünf Rangplätzen erhöhte sich von 1297 auf 1342.

Bei der Simulation *E3* erhielten alle Wortpaare einen unterschiedlichen Wert für den Parameter θ . Für ein Wortpaar wurde der Wert für θ mit Gleichung 8.1 ermittelt. Die Gleichung 8.1 wurde in dem Kapitel 6, Abschnitt Parameterschätzung definiert.

$$\theta_{i,j} = F \cdot (I_i \cdot I_j)^n \quad (8.1)$$

Bei den Simulationen von deutschen Wortassoziationen wurden mit einer Potenzierung von 5 ($n = 5$ in Gleichung 8.1) gute Resultate in Bezug auf die vier Kriterien erzielt. Dies traf aber nicht auf die englischen Wortassoziationen zu. Bei der Voraussage der englischen Wortassoziationen wurde mit einer Potenzierung von 1 ($n = 1$ in Gleichung 8.1) schon die besten Ergebnisse in Bezug auf die vier Kriterien erzielt.

Insgesamt wurden Simulationen mit den Potenzierungen $n = 1$ bis $n = 4$ in Gleichung 8.1 durchgeführt. Die Ergebnisse dieser Simulationen werden im folgenden Abschnitt beschrieben. Für die einzelnen Simulationen mit den unterschiedlichen Potenzierungen musste der optimale Faktor in Gleichung 8.1 bestimmt werden. Bei manchen Faktoren wurden bei einer Fensterlänge von 20 weniger als 460 Primärantworten erzielt. Dies wurde als Kriterium zum Abbruch einer Simulationsreihe gewählt, da bei den vorherigen Simulationen mit einem konstanten θ für alle Wortpaare keine 500 Primärantworten bei größeren Fensterlängen erzielt wurden, wenn die Anzahl richtig vorhergesagter Primärantworten unter 460 bei einer Fensterlänge von 20 lag.

In der Tabelle 8.2 sind die Ergebnisse dargestellt, die mit einer Potenzierung von $n = 1$ erzielt wurden. Mit der Potenzierung von $n = 1$ konnten die Ergebnisse in Bezug auf die vier Kriterien verbessert werden. Die besten Ergebnisse in Bezug auf alle vier Kriterien wurden mit den Faktoren $2.3E - 13$

und $2.4E - 13$ erzielt. Wenn der Faktor $2.4E - 13$ verwendet wurde, wurden die meisten Primärantworten richtig vorausgesagt, dafür wurden aber beim zweiten Kriterium ca. 50 Nennungen der Teilnehmer weniger vorausgesagt als beim Faktor $2.3E - 13$. Bei dem Faktor $2.3E - 13$ wurden aber nur vier Primärantworten weniger vorausgesagt als beim Faktor $2.4E - 13$. Deswegen wurde das Ergebnis, das mit einem Faktor von $2.3E - 13$ erzielt wurde, als das bessere Ergebnis gewertet.

In der nächsten Simulationsreihe sollte untersucht werden, ob eine Potenzierung von $n = 2$ die Ergebnisse in Bezug auf die vier Kriterien noch weiter verbessern könnte.

In der Tabelle 8.3 sind die Ergebnisse, die mit einer Potenzierung von $n = 2$ gewonnen wurden, aufgeführt. Die Ergebnisse in Bezug auf das erste Kriterium hatten sich leicht verschlechtert, während sich die Ergebnisse in Bezug auf das zweite Kriterium etwas verbessert hatten. Die Zahl der richtig vorausgesagten Primärantworten sank bei diesem Faktor aber um 14. Wurden alle vier Kriterien berücksichtigt, so wurde mit einem Faktor von $3.1E - 15$ die besten Ergebnisse erzielt. Nimmt man alle vier Kriterien zusammen, so konnte durch die Potenzierung mit $n = 2$ keine Verbesserung gegenüber dem besten Ergebnis, das bei der vorherigen Simulationsreihe ermittelt wurde, festgestellt werden.

Für die nächste Simulationsreihe wurde eine Potenzierung von $n = 3$ verwendet. Die Ergebnisse werden in der Tabelle 8.4 präsentiert:

Anhand der Tabelle 8.4 ist zu sehen, dass mit einer Potenzierung von $n = 3$ das bisher beste Ergebnis in Bezug auf das zweite Kriterium erzielt wurde. Dieses Ergebnis wurde mit einem Faktor von $1.6E - 17$ bei einer Fensterlänge von 45 erzielt. Die Anzahl richtig vorausgesagter Primärantworten verringerte sich dabei um ca. 20 gegenüber dem besten Ergebnis, das bei einer Potenzierung mit $n = 2$ erzielt wurde. Gegenüber den anderen Kriterien ergab eine Potenzierung mit $n = 3$ keine Verbesserung.

Die Tabelle 8.5 zeigt, dass sich mit einer Potenzierung von $n = 4$ die Ergebnisse in Bezug auf das zweite Kriterium nochmals verbessern ließen. Die Potenzierung mit $n = 4$ führte aber zu einer Verschlechterung der Ergebnisse in Bezug auf das erste und das dritte Kriterium.

Es wurden keine weiteren Potenzierungen mehr durchgeführt, da sich die Ergebnisse in Bezug auf das erste und das dritte Kriterium bei der Potenzierung mit $n = 4$ deutlich verschlechtert hatten. Auch wurden die gute Ergebnisse in Bezug auf die vier Kriterien erst mit einer Fensterlänge von 50 erreicht. Bessere Ergebnisse in Bezug auf die vier Kriterien konnten bei einer Potenzierung von $n = 1$ und $n = 2$ aber schon bei Fensterlängen von 40 und 45 erzielt werden.

Das beste Ergebnis, das mit einem variablen Lernparameter erzielt wurde,

wurde mit folgenden Parameter erreicht:

$$\theta_{i,j} = 2.3E - 13 \cdot I_i \cdot I_j \quad (8.2)$$

Die Anzahl der richtig vorhergesagten Primärantworten konnte von 548, bei der Verwendung eines konstanten Lernparameters, auf 565 erhöht werden. Gleichzeitig erhöhte sich auch die Anzahl aller Nennungen der Teilnehmer ohne die Primärantworten auf den ersten zehn Rangplätzen (Kriterium 2) von 6752 (konstanter Lernparameter: $2E - 6$) auf 7005. Durch die anschließenden Gewichtung der Ergebnisse, die mit einem konstanten Lernparameter erzielt wurden, konnten bei dem zweiten Kriterium sogar 7085 Nennungen gezählt werden. Dafür sank aber die Anzahl richtiger Primärantworten auf 544 ab. Die Ergebnisse in Bezug auf das dritte Kriterium verbesserten sich von 831, bei Verwendung eines konstanten Lernparameters, auf 860. Dies waren drei richtig vorausgesagte Primärantworten weniger als bei der Simulation von *E2*. Beim vierten Kriterium konnte ebenfalls eine Verbesserung von 1297 auf 1363 richtig vorausgesagte Primärantworten beobachtet werden. Bei *E2* wurden 1342 richtig vorausgesagte Primärantworten auf den ersten fünf Rangplätzen gezählt.

Mit dem variablen Lernparameter wurde für 11,3 Prozent der Stimuluswörter die zugehörige Primärantwort richtig vorausgesagt. Für 17,3 Prozent der Stimuluswörter wurde die Primärantwort auf Rang 1 oder Rang 2 richtig vorausgesagt. Für 27,4 Prozent der Stimuluswörter war die Primärantwort der Teilnehmer auf den ersten fünf Rangplätzen zu finden. Pro Stimuluswort wurden 1,4 Responsewörter auf den ersten zehn Rangplätzen vorausgesagt.

Mit dem variablen Lernparameter konnte ein ähnliches Ergebnis in Bezug auf das erste Kriterium erzielt werden, wie Griffiths & Steyvers mit ihrem Topic Modell erzielen konnten. Griffiths & Steyvers konnten mit dem Topic Modell 585 Primärantworten richtig voraussagen (siehe Kapitel 3, Abschnitt Topic Modell). Mit dem Estes Modell wurden etwas weniger Primärantworten vorausgesagt, dies waren 565 Primärantworten. Die erzielten Ergebnisse vom Estes Modell und vom Topic Modell können nicht direkt miteinander verglichen werden, da die Ergebnisse mit unterschiedlichen Textkorpora gewonnen wurden.

Mit einem variablen Lernparameter für die einzelnen Wortpaare wurde schon ohne Potenzierung ($n = 1$) das beste Ergebnis in Bezug auf die vier Kriterien erreicht. Dagegen wurde bei der Berechnung der deutschen Wortassoziationen eine Potenzierung mit $n = 5$ benötigt, um das beste Ergebnis in Bezug auf die vier Kriterien zu erhalten. Dies mag daran gelegen haben, dass bei der Berechnung der deutschen Wortassoziationen vor der Berech-

nung eine Lemmatisierung vorgenommen wurde. Ein weiterer Unterschied, der sich bei der Simulation mit englischen Wortassoziationen ergab, war die Fensterlänge, bei der das beste Ergebnis in Bezug auf die vier Kriterien ermittelt wurde. Bei den deutschen Wortassoziationen erhielt man schon ab einer Fensterlänge von 22 gute Ergebnisse in Bezug auf die vier Kriterien, während bei den englischen Wortassoziationen erst ab einer Fensterlänge von 40 gute Ergebnisse in Bezug auf die vier Kriterien erzielt werden konnten.

Fazit: Mit dem Estes Modell konnten auch englische Wortassoziationen vorausgesagt werden. Das erzielte Ergebnis in Bezug auf das erste Kriterium unterschied sich nur wenig von dem Ergebnis, das Griffiths & Steyvers mit ihrem Topic Modell erzielt hatten. Für die Voraussage von englischen Wortassoziationen wurde keine Lemmatisierung benötigt.

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$1.6E - 12$	356	4204	563	868	20
$1.6E - 13$	505	6371	779	1222	50
$1.9E - 13$	549	6774	834	1312	45
$2.1E - 13$	560	6930	858	1336	50
$2.2E - 13$	561	6951	855	1336	50
$2.3E - 13$	565	7005	860	1363	45
$2.4E - 13$	569	6951	866	1354	40
$2.5E - 13$	560	6961	859	1361	45
$2.6E - 13$	549	6946	861	1348	45
$3.6E - 13$	529	6250	809	1241	50
$4.6E - 13$	557	6627	852	1312	50
$5.6E - 13$	552	6559	850	1295	50
$6.6E - 13$	532	5967	814	1238	45
$1.6E - 14$	284	4191	422	716	20

Tabelle 8.2: Simulation von englischen Wortassoziationen mit dem Estes Modell, wobei ein variabler Wert für θ verwendet wurde, $\theta_{i,j} = F \cdot (I_i \cdot I_j)$.

Es sind die Ergebnisse in Bezug auf die vier Kriterien in Abhängigkeit von den unterschiedlichen Faktoren dargestellt. Bei den Faktoren $1.6E - 12$ und $1.6E - 14$ wurden bei einer Fensterlänge von 20 weniger als 460 richtig vorhergesagte Primärantworten erzielt. Deswegen wurden mit diesen beiden Faktoren keine weiteren Simulationen mehr durchgeführt.

Die meisten Primärantworten der Teilnehmer wurden mit einem Faktor von $2.4E - 13$ bei einer Fensterlänge von 40 richtig vorausgesagt, und zwar waren dies 569 Primärantworten. Das beste Ergebnis in Bezug auf das dritte Kriterium wurde ebenfalls mit diesem Faktor erzielt. Dies waren 866 Primärantworten auf den ersten beiden Rangplätzen.

Das beste Ergebnis in Bezug auf das zweite und das vierte Kriterium wurde mit einem Faktor von $2.3E - 13$ bei einer Fensterlänge von 45 berechnet, dabei sank die Zahl der richtig vorausgesagten Primärantworten um 4. Bei diesem Faktor wurden 7005 Nennungen aller Teilnehmer ohne die Primärantwort auf den ersten zehn Rangplätzen, und 1363 Primärantworten der Teilnehmer auf den ersten fünf Rangplätzen richtig vorausgesagt.

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$1.6E - 13$	289	3049	460	700	20
$1.6E - 14$	412	5137	638	1039	20
$1.6E - 15$	537	7068	832	1336	35
$1.9E - 15$	540	6706	823	1286	50
$2.1E - 15$	536	6724	836	1307	50
$2.3E - 15$	540	6870	846	1319	50
$2.6E - 15$	548	6944	854	1327	45
$2.9E - 15$	544	6930	857	1332	45
$3.1E - 15$	551	6922	871	1338	45
$3.2E - 15$	549	6886	872	1321	45
$3.4E - 15$	548	6847	867	1325	45
$3.5E - 15$	547	6828	865	1332	45
$3.6E - 15$	557	6817	868	1328	45
$3.8E - 15$	556	6722	865	1330	50
$4.1E - 15$	550	6657	849	1323	50
$4.6E - 15$	535	6473	844	1305	45
$5.6E - 15$	537	6463	834	1288	50
$1.6E - 16$	412	5453	614	1031	20

Tabelle 8.3: Simulation von englischen Wortassoziationen mit dem Estes Modell mit dem Parameter $\theta_{i,j} = F \cdot (I_i \cdot I_j)^2$.

Die Tabelle zeigt die erzielten Ergebnisse in Bezug auf die vier Kriterien in Abhängigkeit von den unterschiedlichen Faktoren. Mit einem Faktor von $3.6E - 15$ konnten die meisten Primärantworten (557) bei einer Fensterlänge von 45 richtig vorausgesagt werden.

Bei dem zweiten Kriterium wurden bei einem Faktor von $1.6E - 15$ die besten Resultate schon bei einer Fensterlänge von 35 erzielt. Die besten Ergebnisse in Bezug auf das dritte und das vierte Kriterium wurden ebenfalls bei einem Faktor von $3.1E - 15$ erzielt.

Die Simulationsreihen, die in der Spalte Fensterlänge eine 20 haben, wurden abgebrochen, da sie bei dieser Fensterlänge nicht mindestens 460 Primärantworten richtig voraussagen konnten.

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$1.6E - 16$	453	6411	713	1188	20
$6.6E - 16$	426	5622	675	1108	20
$8E - 16$	424	5558	673	1095	20
$1.3E - 17$	510	7020	824	1334	50
$1.6E - 17$	530	7104	842	1343	45
$1.7E - 17$	524	7093	842	1350	50
$1.9E - 17$	535	7071	846	1347	50
$2.1E - 17$	534	7096	849	1344	45
$2.3E - 17$	535	7062	856	1337	50
$2.6E - 17$	527	7034	843	1351	50
$2.9E - 17$	529	7015	843	1349	45
$3.1E - 17$	540	6959	844	1321	50
$3.3E - 17$	537	6983	858	1335	45
$3.6E - 17$	528	6924	855	1322	50
$1.6E - 18$	443	6729	672	1128	20

Tabelle 8.4: Simulation von englischen Wortassoziationen mit dem Estes Modell mit dem Parameter $\theta_{i,j} = F \cdot (I_i \cdot I_j)^3$.

Die Tabelle zeigt die Ergebnisse in Bezug auf die vier Kriterien in Abhängigkeit von den unterschiedlichen Faktoren. Bei einem Faktor von $1.6E - 17$ wurde das beste Ergebnis in Bezug auf das zweite Kriterium erzielt. Das beste Ergebnis in Bezug auf das erste Kriterium wurde bei einem Faktor von $3.1E - 17$ erzielt.

Die Simulationsreihen, die in der Spalte Fensterlänge eine 20 haben, wurden abgebrochen, da sie bei dieser Fensterlänge nicht mindestens 460 Primärantworten richtig voraussagen konnten.

Faktor F	Krit. 1	Krit. 2	Krit. 3	Krit. 4	Fensterlänge
$1.6E - 16$	449	6870	726	1200	20
$1.6E - 17$	444	6834	727	1222	20
$1.6E - 18$	456	7028	734	1236	20
$3.6E - 18$	414	6903	692	1155	20
$1.6E - 19$	499	7190	807	1337	45
$2.6E - 19$	499	7144	811	1321	45
$3.6E - 19$	503	7212	825	1342	50
$3.8E - 19$	502	7204	810	1342	45
$4.1E - 19$	499	7210	829	1337	50
$4.4E - 19$	506	7194	825	1343	50
$4.6E - 19$	513	7191	827	1345	50
$4.9E - 19$	504	7209	824	1329	45
$5.6E - 19$	498	7198	822	1343	50
$1.6E - 20$	437	6853	689	1152	20

Tabelle 8.5:

Simulation von englischen Wortassoziationen mit dem Estes Modell mit dem Parameter $\theta_{i,j} = F \cdot (I_i) \cdot I_j^4$.

Die Tabelle zeigt die Ergebnisse in Bezug auf die vier Kriterien in Abhängigkeit von den unterschiedlichen Faktoren. Bei einem Faktor von $3.6E - 19$ wurde das beste Ergebnis in Bezug auf das zweite Kriterium erzielt. Das beste Ergebnis in Bezug auf das erste Kriterium wurde bei einem Faktor von $4.6E - 19$ erzielt.

Das beste Ergebnis in Bezug auf alle vier Kriterien wurde mit einem Faktor $4.6E - 19$ bei einer Fensterlänge von 50 erzielt. Dies waren 513 richtig vorausgesagte Primärantworten, 7191 Nennungen aller Teilnehmer ohne die Primärantworten auf den ersten zehn Rangplätzen, 827 richtig vorausgesagte Primärantworten auf den ersten beiden Rangplätzen und 1345 richtig vorausgesagte Primärantworten auf den ersten fünf Rangplätzen.

Die Simulationsreihen, die in der Spalte Fensterlänge eine 20 haben, wurden abgebrochen, da sie bei dieser Fensterlänge nicht mindestens 460 Primärantworten richtig voraussagen konnten.

Kapitel 9

Zusammenfassung und Interpretation der Ergebnisse

In dieser Arbeit sollte untersucht werden, wie Wortassoziation entstehen können. Die Hypothese war: Wortassoziationen sind die Ergebnisse von Lernprozessen und können somit mit Hilfe von mathematischen Lerntheorien erklärt werden. Zur Überprüfung dieser Hypothese wurden drei mathematische Lerntheorien ausgewählt, die Stimulusauswahltheorie von Estes, die Theorie zur Schätzung der Auftretensrate (Rate Estimation Theorie) von Gallistel und die Theorie von Rescorla & Wagner. Alle drei Theorien enthalten Lernmodelle, die zur Simulation des Erlernens von Wortassoziationen verwendet werden konnten. Mit allen drei Lernmodellen, Estes Modell, Rescorla & Wagner Modell und Gallistel Modell, war es gelungen, erfolgreich das Erlernen von deutschen Wortassoziationen mit dem Computer zu simulieren. Die berechneten Assoziationen wurden anhand der Assoziationsnormen, die von Russell & Meseck erhoben wurden, überprüft. Die drei Modelle konnten ungefähr dieselbe Anzahl von Primärantworten, die im Mittel ein Teilnehmer in der Erhebung von Russell & Meseck produziert hatte, voraussagen.

Das Estes Modell gehört zu der Klasse der Kontiguitätsmodelle, während das Rescorla & Wagner Modell zur Klasse der Kontingenztmodelle gehört. Da sich beide Modelle zur Voraussage von Wortassoziationen eignen, reicht für die Voraussage von Wortassoziationen offenbar ein Kontiguitätsmodell aus.

Wettler und Rapp (Wettler, Rapp & Sedlmeier, 2005) konnten das Erlernen von Wortassoziationen bereits mit Hilfe eines Kontiguitätsmodells erklären. Im Unterschied zu dieser Arbeit wurden die Wortassoziationen bei Wettler und Rapp auf der Basis von Kookkurrenzen ermittelt. Für diese Arbeit wurden keine Kookkurrenzen verwendet, sondern die Simulationen erfolgten durch die inkrementelle Anwendung von Lern- und Interferenzgleichungen. Der Arbeit von Wettler und Rapp lagen die gleichen Lern- und

Interferenzgleichungen zugrunde. Sie verwendeten aber statt der Lern- und Interferenzgleichungen die bedingte Wahrscheinlichkeit, die sich aber nur bei stochastische Unabhängigkeit aus den Lern- und Interferenzgleichungen herleiten lässt, zur Berechnung von Wortassoziationen. Das Auftreten von Wörtern ist aber nicht stochastisch unabhängig voneinander.

Der Vorteil bei der Verwendung von bedingten Wahrscheinlichkeiten zur Berechnung von Wortassoziationen war, dass nur die Kookkurrenzen und die Einzelhäufigkeiten der Wörter benötigt wurden. Durch die Verwendung von Kookkurrenzen konnten Wettler und Rapp bei der Assoziationsberechnung ein viel größeres Vokabular verwenden. Ein so großes Vokabular konnte aus Laufzeit- und Speicherplatzgründen und aufgrund der vielen Simulationsreihen, die zur Parameterschätzung benötigt wurden, bei den inkrementellen Lernmodellen nicht verwendet werden.

Die höhere Laufzeit gegenüber dem Modell von Wettler und Rapp entstand einmal durch die Verwendung von Gleitkommazahlen und dadurch, dass wesentlich mehr Rechenoperationen durch die Verwendung einer Interferenzregel entstanden. Wettler und Rapp mussten nur die gemeinsame Auftretenshäufigkeit von Wortpaaren zählen. Bei dem Estes und dem Rescorla & Wagner Modell wurde beim gemeinsamen Auftreten von zwei Wörtern eine Lerngleichung ausgeführt, die Gleitkommazahlen verwendete. Trat jeweils nur ein Wort aus einem Wortpaar auf, so musste bei dem Modell von Wettler und Rapp keine Operation durchgeführt werden, während das Estes und das Rescorla & Wagner Modell die Anwendung einer Interferenzregel vorsahen.

Beim Gallistel Modell wurden beim Einlesen des Textkorpus nur Häufigkeiten gezählt. Deswegen benötigte dieses Modell auch nicht so viel Rechenzeit wie die anderen beiden Modelle. Dafür wurde aber mehr Speicherplatz benötigt, da nicht nur eine Matrix für die Assoziationsstärken benötigt wurde. Es wurde eine Matrix für die Zeiten und eine Matrix für die Auftretenshäufigkeiten benötigt. Da beim Gallistel Modell ein lineares Gleichungssystem gelöst werden musste, musste nach dem Einlesen des Textkorpus noch die inverse Matrix berechnet werden, die noch zusätzlich zu den beiden anderen Matrizen Speicherplatz benötigte.

Wegen der aufgeführten Probleme mit der Rechenzeit und des Speicherplatzes wurde für die Simulation der deutschen Wortassoziationen nur ein Vokabular von 802 Wörtern verwendet, während Rapp (Wettler, Rapp & Ferber, 1993) ein Vokabular von 65356 Wörtern verwenden konnte. Die geringe Vokabulargröße bedeutet eine Einschränkung der Allgemeingültigkeit der erzielten Resultate. Deshalb wurde zur Simulation der englischen Wortassoziationen ein größeres Vokabular verwendet. Dieses Vokabular enthielt 9126 Wörter. Wegen dem größeren Vokabular und dem daraus resultierenden Rechenaufwand konnten die Simulationen der englischen Wortassozia-

tionen aber nur mit einem Lernmodell durchgeführt werden. Dies war das Estes Modell, da dieses Modell leicht bessere Ergebnisse bei der Voraussage von deutschen Wortassoziationen erzielt hatte als die anderen Modelle. Aber auch mit diesem großen Vokabular konnten gute Ergebnisse erzielt werden. Es konnten 565 Primärantworten richtig vorausgesagt werden. Das Topic Modell von Griffiths und Steyvers verwendete die gleichen Assoziationsnormen aber einen anderen Textkorpus. Mit ihrem Modell konnten Griffiths und Steyvers 585 Primärantworten (Griffiths & Steyvers, 2002) richtig voraussagen. Dies waren etwas mehr als beim Estes Modell. Da aber unterschiedliche Korpora zur Berechnung der Wortassoziationen verwendet wurden, konnten diese Ergebnisse nicht direkt miteinander verglichen werden, sie gaben nur einen ungefähren Richtwert an.

Ein Worthäufigkeitseffekt trat bei allen drei Modellen auf. Am stärksten wurde dieser bei dem Gallistel Modell beobachtet. Durch eine anschließende Gewichtung der berechneten Ergebnisse konnte der Worthäufigkeitseffekt aber stark verringert werden. Bei den anderen beiden Modellen wurde der Worthäufigkeitseffekt durch eine geeignete Wahl der freien Parameter verringert. Für die Reduzierung des Worthäufigkeitseffekts wurde der Informationsgehalt eines Wortes verwendet. Dieser reichte alleine aber nicht zur Reduzierung des Worthäufigkeitseffekts aus, sondern der Informationsgehalt musste noch potenziert werden. Auffallend war, dass bei allen drei Modellen eine Potenzierung zwischen 5 und 6 einschließlich die besten Resultate bei der Simulation der deutschen Wortassoziationen lieferte. Bei der Simulation der englischen Wortassoziationen, die nur mit dem Estes Modell durchgeführt wurden, wurde nur der Informationsgehalt des Wortes verwendet. Dieser brauchte nicht potenziert zu werden. Dies mag einmal daran gelegen haben, dass für die Simulation der englischen Wortassoziationen keine Lemmatisierung des Textkorpus vorgenommen wurde, oder an der englischen Sprache.

Die Lemmatisierung des Textkorpus führte zu einer Erhöhung der Worthäufigkeiten. Insbesondere bei mehrdeutigen Wörtern konnte durch die Lemmatisierung die Worthäufigkeit stark zunehmen, z.B. beim Wort *sein*. Ob auch bei der Berechnung der deutschen Assoziationen auf eine Lemmatisierung verzichtet werden kann, wäre eine weitere Untersuchung wert. Bei dieser Untersuchung könnte auch geklärt werden, ob ohne Lemmatisierung auch eine Potenzierung zwischen 5 und 6 nötig ist, um den Worthäufigkeitseffekt zu korrigieren. Bei den englischen Assoziationsberechnungen könnte überprüft werden, ob eine Lemmatisierung bei den englischen Wörtern zu besseren Ergebnissen führen würde.

Literaturverzeichnis

- Berry M. W., Dumais S. T. und O'Brien G. W. (1995). Using Linear Algebra for Intelligent Information Retrieval. *SIAM:Review*, 37,573-595.
- Böhnisch M. (1999). *Vorhersage menschlicher Wortassoziationen*. Diplomarbeit.
- Bower G. H. (1994). A Turning Point in Mathematical Learning Theory. *Psychological Review*, Volume 101 Number 2,290-300.
- Brown R. & Berko J. (1960). Word association and the acquisition of grammar. *Child Development*, 31, 1-14.
- Burgess C. (1998). From simple associations to the building blocks of language: Modeling meaning in memory with the HAL model. *Behavior Research Methods, Instrumentation, and Computers*, 30 (2), 188-198.
- Chapman G. B. & Robbins S. I. (1990). Cue interaction in contingency judgement. *Memory and Cognition*, 18, 537-545.
- Cheng P. W. (1997). From Covariation to Causation: A Causal Power Theory. *Psychological Review*, 1997, Vol. 104, No. 2,367-405.
- Church K.W. (1988). A stochastic parts program and noun phrase parser for unrestricted Text. *Second Conference on Applied Natural Language Processing*, Austin, Texas, 136-143.
- Clark H. H. (1970). Word association and linguistic theory. In J. Lyons (Ed.), *New horizons in linguistic*. Harmondsworth: Penguin, 271-286.
- Coombs C. H., Dawes R. H. & Tversky A. (1970). *Mathematical Psychology, An Elementary Introduction*. Prentice-Hall: New Jersey.
- Danks D. (2003). Equilibria of the Rescorla & Wagner model. *Journal of Mathematical Psychology*, 47,109-121.

- Deerwester S., Dumais S. T., Furnas G. W., Landauer T. K. & Harshmani R. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society of Information Science*, 41, 391-407.
- Deese J. (1959b). On the prediction of occurrences of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 58, 17-22.
- Estes W. K. (1950). Toward a Statistical Theory of Learning. *Psychological Review*, 57, 94-107.
- Ferber R., Wettler M. & Rapp R. (1995). An associative model of word selection in the generation of search queries. *Journal of the American Society for Information Science*, 46, 685-699.
- Foppa K. (1965). *Lernen Gedächtnis Verhalten - Ergebnisse und Probleme der Lernpsychologie*. Verlag Kiepenheuer und Witsch.
- Gallistel C. R. (1990). *The Organization of Learning*. MIT Press.
- Gallistel C. R. & Gibbon J. (2000). Time, Rate and Conditioning. *Psychological Review* 107, 289-334.
- Galton F. (1880). Psychometric experiments. *Brain*, 2, 149-162.
- Griffiths T. L. & Steyvers M. (2002). Prediction and Semantic Association. *Advances in Neural Information Processing Systems*, 15, pp . 11-18.
- Hilgard E. R. & Bower G. H. (1983). *Theorie des Lernens I*. Stuttgart: Ernst Klett Verlag.
- Hofmann, T. (1999). Probabilistic Latent Semantic Analysis. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*.
- Healy A. F., Kosslyn S. M. & Shiffrin R. M. (1992). *From learning theory to connectionist theory: Essays in honor of William K. Estes*. Vol 1 & 2, Hillsdale, NJ:Erlbaum.
- James W. (1890). *The principles of psychology*. New York: Dover Publication.
- Kohonen T. (1989). *Self-organiz ation and associative memory*. Berlin: Springer.
- Landauer T. (2002). On the computational basis of learning and cognition: Arguments from LSA, aus B.H Ross (Ed), *The Psychology of Learning and Motivation*. 41, 43-84.

- Landauer T. K. & Dumais S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Aquisition, Induction, and Representation of Knowledge. *Psychological Review*, Vol. 104, No. 2, 211-240.
- Landauer T. K., Foltz P. W. & Laham D. (1998). An Introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259-284.
- Lashley K. S. (1951). The problem of serial order in behavior. In L.A. Jeffres (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley.
- Lay D. C. (1994). *Linear Algebra and its Applications*. Addison-Wesley Publishing Company.
- Lemaire B. & Denhiere G. (2004). Incremental Construction of an Associative Network from a Corpus. In Forbus, Kenneth, Gentner, Dedre, Regier & Terry (Eds.), *Proceedings 26th Annual Meeting of the Cognitive Science Society* pp. 825-830. Chicago.
- Lezius W., Rapp R. & Wettler M. (1998). Freely Available Morphological Analyzer, Disambiguator, and Context Sensitive Lemmatizer for German. In: *Proceedings of the COLING-ACL*.
- Lezius W., Rapp R. & Wettler M. (1996). A Morphology-System and Part-of-Speech Tagger for German. In: D. Gibbon, (Eds.), *Natural Language Processing and Speech Technology. Results of the 3rd KONVENS Conference* (pp. 369-278), Mouton de Gruyter, Berlin, 1996.
- Lund K. & Burgess C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instrumentation, and Computers*, 28, 203-208.
- Matthäus W. (1980). Theoretische und experimentelle Untersuchungen zum verbalen Assoziieren. Methodenkritisch gerichtete Strukturanalysen von Einfallsreihen und ihrer Veränderung bei Aufgabenwechsel. *Bochumer pädagogische und psychologische Studien Nr. 5*. Bochum: Studienverlag Dr. N. Brockmeyer.
- Michalewicz Z. & Fogel D. B. (2000). *How to Solve It: Modern Heuristics*. Springer Verlag.
- Nelson D. L., McEvoy C. L. & T. A. Schreiber (1999). *The University of South Florida word association norms*, <http://www.usf.edu/FreeAssociation>.
- Murdock B. B. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.

- Raaijmakers J.G.W. & Shiffrin R.M. (1980). SAM: A theory of probabilistic search of associative memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*, Vol. 14, pp. 207-262.
- Rayner K. & Pollatsek A. (1989). *The Psychology of Reading*. Lawrence Erlbaum Associates: New Jersey.
- Palermo D. & Jenkins J. (1964). *Word association norms*. Minneapolis, MN: University of Minnesota Press.
- Rapp R. (1996). *Die Berechnung von Assoziationen: ein korpuslinguistischer Ansatz*. Hildesheim, Zürich, New York
- Razran G. A. (1939). Quantitative study of meaning by conditioned salivary technique (semantic conditioning). *Science*, 90, 89-91.
- Russell W. A. (1970). The complete german language norms for responses to 100 words from the Kent-Rosanoff Word Association Test. In L. Postmann & G. Keppel (Eds.), *Norms of word association* (pp. 53-94). New York: Academic Press.
- Russell W. A. & Meseck O. R. (1959). Der Einfluß der Assoziation auf das Erinnern von Worten in der deutschen, französischen und englischen Sprache. *Zeitschrift für experimentelle und angewandte Psychologie*, 6, 191-211.
- Sarnoff A. M. (1962). The associative basis of the creative process. *Psychological Review*, Vol. 69, No. 3, 220-232.
- Schwartz B. & Reisberg D. (1991). *Learning and Memory*. New York: Norton.
- Sedlmeier P. (1999). *Improving statistical reasoning: Theoretical models and practical implications*. Mahwah, NJ: Erlbaum.
- Staats A. W. (1968). *Learning, Language, and Cognition*. New York: Holt, Rinehart & Winston.
- Sutton B., Richard S. & Andrew G. (1981). Toward a Modern Theory of Adaptive Networks: Expectation and Prediction. *Psychological Review*, 88, 135-170.
- Spence D. P. & Owens K. C. (1990). Lexical Co-Occurrence and Association Strength. *Journal of Psycholinguistic Research*, Vol. 19, No. 5, pp. 317-330.

- Steyvers M. & Griffiths T. (in press). Probabilistic Topics Models. In T. Landauer, D McNamara, S. Dennis & W. Kintsch (Eds.), *Latent Semantic Analysis: A Road to Meaning*. Laurence Erlbaum.
- Wagner A. R. & Rescorla R. A. (1972). Inhibition in Pavlovian conditioning: Applications of a theory. In R. A. Boakes & S. Halliday (Eds.), *Inhibition and learning* (pp. 301-36). New York: Academic Press.
- Wettler M. (1980). *Sprache Gedächtnis Verstehen*. de Gruyter Studienbuch: Grundlagen der Kommunikation.
- Wettler M. (2002). Free Word Associations and the Frequency of Co-occurrence in Language use. In: Sedlmeier, P. & Betsch, T. (Eds.), *Frequency Processing and Cognition*, pp. 271-284. Oxford UK: Oxford University Press.
- Wettler M., Rapp R. & Ferber R. (1993). Freie Assoziationen und Kontiguitäten von Wörtern in Texten. *Zeitschrift für Psychologie*, 201, 99-108.
- Wettler M., Rapp R. & Sedlmeier P. (2005). Free Word Associations Correspond to Contiguities between Words in Texts. *Journal of Quantitative Linguistics*, Vol. 12, No. 2, pp. 111-122.
- Wahrig G. (1997). *Deutsches Wörterbuch*, Bertelsmann.