

Universität Bielefeld
Fakultät für Linguistik und Literaturwissenschaft

**Evaluation von Wortgenerierungsleistungen
zur Diagnose
kommunikativ-kognitiver Defizite**

DISSERTATION ZUR ERLANGUNG DER DOKTORWÜRDE
IM FACH KLINISCHE LINGUISTIK

vorgelegt von
Kristina Thiele
Mai 2013

Gutachter: Prof. Dr. Prisca Stenneken
Dr. Joana Cholin

Gedruckt auf alterungsbeständigem Papier °° ISO 9706

Dank

Mein besonderer Dank gilt Frau Prof. Dr. Prisca Stenneken, die mich als meine Betreuerin in meinem Tun stets unterstützt, meine Ideen gefördert und durch konstruktive Kritik bereichert hat. Sie hat stets Zeit für mich, meine Fragen und die Betreuung meines Promotionsprojektes gefunden und dadurch maßgeblich zum Gelingen dieser Arbeit beigetragen.

Weiterhin möchte ich mich bei Frau Dr. Joana Cholin bedanken: für die Begutachtung dieser Dissertation sowie die stets offenen Gespräche und wertvollen Tipps für die Arbeit im Wissenschaftsdschungel.

Herzlichst danken möchte ich auch Karoline Malchus für die allmorgendlichen motivierenden Skype-Konferenzen, Dr. Katrin Johannsen und Dr. Petra Jaecks für ihr jederzeit offenes Ohr und ihr Engagement bei der Besprechung der verschiedensten Themen und Fragen sowie Maria Trüggelmann für die angenehme Zeit im gemeinsamen Büro.

Ein besonderer Dank gilt auch dem SFB 673 *Alignment in Communication* sowie dem Exzellenzcluster *Cognitive Interaction Technology*, die mein Promotionsprojekt nicht nur finanziell unterstützt, sondern auch durch die Möglichkeit zum interdisziplinären Austausch zum Gelingen dieser Arbeit beigetragen haben.

Ganz besonders möchte ich mich auch bei allen Teilnehmern meiner Studien für ihren Einsatz bei der Wortgenerierung bedanken.

Weiterhin danke ich meiner Familie und meinen Freunden, durch die ich stets sehr viel Unterstützung erfahren habe. Ein besonderer Dank geht an meine Schwester Leonie, für ihren unermüdlichen Einsatz beim Korrekturlesen.

Schließlich möchte ich mich bei Robin Jonas bedanken, der mich immer wieder motiviert hat weiter zu machen!

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Tabellenverzeichnis	IV
Abbildungsverzeichnis	VII
1 Einleitung	1
1.1 Motivation & Ziele	2
1.2 Aufbau & Inhalt	3
2 Wortgenerierung im klinisch-linguistischen Kontext	5
2.1 Modelltheoretische Perspektiven	7
2.1.1 Mannheimer Regulationstheorie der Sprachverarbeitung (Herrmann & Grabowski, 1994)	8
2.2 Exekutivfunktionen	11
2.2.1 Definition	11
2.2.2 Neuroanatomische Korrelate	11
2.2.3 Modelle & Taxonomie	14
2.3 Wortgenerierungsfähigkeit	18
2.3.1 Aufgabentypen	19
2.3.2 Funktionen & neuroanatomische Korrelate	22
2.4 Sprach- und Kommunikationsstörungen	25
2.4.1 Sprachsystematische Defizite: Aphasie	26
2.4.2 Kommunikativ-kognitive Defizite	30
2.4.3 Beeinträchtigungen von Sprache und Kommunikation bei Multipler Sklerose	34
2.5 Diagnostik	41
2.5.1 Evaluation von Wortgenerierungsleistungen	45
2.6 <i>machine learning</i>	48
2.6.1 <i>machine learning</i> in der klinischen Anwendung	51
2.7 Fragestellungen	53
3 Evaluation quantitativer und qualitativer Charakteristika der Wortgenerierungsleistung bei Multipler Sklerose	54
3.1 Einleitung & Fragestellungen	54
3.2 Methode	58

INHALTSVERZEICHNIS

3.2.1	Versuchsteilnehmer	59
3.2.2	Material, Durchführung & Auswertung	61
3.3	Ergebnisse	64
3.3.1	Wortgenerierungsleistungen	65
3.3.2	Anteil korrekter Nennungen & Evaluation der Fehler- typen	70
3.4	Diskussion	76
3.4.1	Interpretation der Wortgenerierungsleistungen	77
3.4.2	Interpretation des Anteils korrekter Nennungen sowie der auftretenden Fehler	80
3.4.3	Fazit	82
4	Evaluation von Wortgenerierungsleistungen zur Klassifikation von Sprach- und Kommunikationsstörungen mit <i>machine learning</i>-Techniken	84
4.1	Einleitung & Fragestellung	84
4.2	Methode	87
4.2.1	Versuchsteilnehmer	87
4.2.2	Material & Durchführung	89
4.3	Ergebnisse	93
4.3.1	Klassifikation	94
4.3.2	Variablenanalyse	100
4.4	Diskussion	103
4.4.1	Fazit	107
5	Exploration einer Methodologie zur Evaluation von Wort- generierungsleistungen mit <i>machine learning</i>-Techniken	108
5.1	Einleitung & Fragestellung	108
5.2	Methode (allgemein)	111
5.2.1	Versuchsteilnehmer	111
5.2.2	Material, Durchführung & Auswertung	112
5.2.3	Evaluation mit <i>machine learning</i> -Techniken	116
5.3	Experiment 1	117
5.3.1	Experiment 1a	117
5.3.2	Experiment 1b	120
5.4	Experiment 2	126
5.4.1	Experiment 2a	126
5.4.2	Experiment 2b	127
5.5	Experiment 3	130
5.6	Diskussion	134
5.6.1	Fazit	140

6	Diskussion	141
6.1	Zusammenfassung	142
6.2	Evaluation von Wortgenerierungsleistungen	145
6.2.1	Evaluation quantitativer und qualitativer Auswertungsparameter: Zur Bedeutung zeitlicher Charakteristika	146
6.2.2	Evaluation von Leistungen in unterschiedlichen Wortgenerierungsaufgaben	151
6.2.3	Sozial-kognitive Charakteristika von Wortgenerierungsleistungen: Ein Einzelfall	156
6.2.4	Evaluation von Wortgenerierungsleistungen mit <i>machine learning</i> -Techniken	161
6.3	Fazit	164
6.4	Ausblick	168
	Literaturverzeichnis	171

Tabellenverzeichnis

2.1	<i>Beispiele simpler lexikalischer Wortgenerierungsaufgaben . . .</i>	20
2.2	<i>Beispiele simpler semantischer Wortgenerierungsaufgaben . .</i>	20
2.3	<i>Beispiele intradimensionaler lexikalischer und semantischer Wechsel-Aufgaben</i>	21
3.1	<i>Übersicht über die demographischen Daten (Versuchsteilnehmer (VT), Alter, Geschlecht (m/w), Krankheitsdauer (KHD) in Monaten) sowie Test-Ergebnisse aus dem DemTect (Kalbe et al., 2004; Test-Score: 13 bis 18 = altersentsprechende Leistung; 9 bis 12 = leichte kognitive Beeinträchtigung; 0 bis 8 = Verdacht auf Demenz; k.A. = keine Angabe) und Angaben zum Sprechtempo (MVP-ST) aus dem Münchener Verständlichkeitsprofil (Ziegler & Zierdt, 2008)</i>	61
3.2	<i>Beispiele möglicher Fehlertypen in den lexikalischen und semantischen Wortgenerierungsaufgaben, Fehler sind mit * markiert</i>	63
3.3	<i>Kategorisierung der Fehlertypen in der Verbgenerierungsaufgabe, Fehler sind mit * markiert</i>	64
4.1	<i>Accuracy gemittelt über alle Aufgaben (Patienten mit Aphasie vs. Sprachgesunde)</i>	94
4.2	<i>Accuracy gemittelt über alle Klassifikatoren (Patienten mit Aphasie vs. Sprachgesunde)</i>	95
4.3	<i>Accuracy gemittelt über die simplen lexikalischen Aufgaben (Patienten mit Aphasie vs. Sprachgesunde)</i>	95
4.4	<i>Accuracy in der simplen semantischen Aufgabe (Patienten mit Aphasie vs. Sprachgesunde)</i>	96
4.5	<i>Accuracy gemittelt über die Wechsel-Aufgaben (Patienten mit Aphasie vs. Sprachgesunde)</i>	96
4.6	<i>Accuracy gemittelt über alle Aufgaben (Patienten mit KKD vs. Sprachgesunde)</i>	97
4.7	<i>Accuracy gemittelt über alle Klassifikatoren (Patienten mit KKD vs. Sprachgesunde)</i>	98

4.8	<i>Accuracy gemittelt über die simplen lexikalischen Aufgaben (Patienten mit KKD vs. Sprachgesunde)</i>	98
4.9	<i>Accuracy in der simplen semantischen Aufgabe (Patienten mit KKD vs. Sprachgesunde)</i>	98
4.10	<i>Accuracy gemittelt über die Wechsel-Aufgaben (Patienten mit KKD vs. Sprachgesunde)</i>	99
4.11	<i>Accuracy gemittelt über alle Klassifikatoren (Patienten mit KKD vs. Patienten mit Aphasie)</i>	99
4.12	<i>Accuracy gemittelt über alle Aufgaben (Patienten mit KKD vs. Patienten mit Aphasie)</i>	100
4.13	<i>Liste der besten zehn Variablen gemittelt über alle Aufgaben (Patienten mit Aphasie vs. Sprachgesunde)</i>	101
4.14	<i>Liste der besten zehn Variablen gemittelt über alle Aufgaben (Patienten mit Aphasie vs. Patienten mit KKD)</i>	102
4.15	<i>Liste der besten Variablen gemittelt über alle Aufgaben (Patienten mit KKD vs. Sprachgesunde)</i>	103
5.1	<i>Liste der besten zehn Variablen (gemittelt über alle Aufgaben) - Experiment 1a</i>	118
5.2	<i>Mittelwerte und Ergebnisse des t-Tests unabhängiger Stichproben für die Variablen mit der besten accuracy (gemittelt über alle Aufgaben)</i>	120
5.3	<i>Liste der Variablen mit der besten accuracy in den vier Wortgenerierungsaufgaben - Experiment 1b</i>	122
5.4	<i>Mittelwerte und Ergebnisse des t-Tests unabhängiger Stichproben für die Baseline-Variable nr_responses in den vier Wortgenerierungsaufgaben</i>	123
5.5	<i>Einschätzung der Wortgenerierungsleistung im Vergleich zur Referenzstichprobe aus dem Regensburger Wortflüssigkeitstest (Aschenbrenner et al., 2000) in den vier Wortgenerierungsaufgaben für Patienten mit KKD und Sprachgesunde; Angegeben sind jeweils die Häufigkeiten unterdurchschnittlicher (ULei), durchschnittlicher (DLei) sowie überdurchschnittlicher Leistung (ÜLei) in den beiden Gruppen</i>	123
5.6	<i>Mittelwerte und Ergebnisse des t-Tests unabhängiger Stichproben für die Variablen mit der besten accuracy in Aufgabe LEX_M</i>	124
5.7	<i>Mittelwerte und Ergebnisse des t-Tests unabhängiger Stichproben für die Variablen mit der besten accuracy in Aufgabe SEM_Tiere</i>	125
5.8	<i>Mittelwerte und Ergebnisse des t-Tests unabhängiger Stichproben für die Variablen mit der besten accuracy in Aufgabe LEX_H/T</i>	125

TABELLENVERZEICHNIS

5.9	<i>Mittelwerte und Ergebnisse des t-Tests unabhängiger Stichproben für die Variablen mit der besten accuracy in Aufgabe SEM_Kleidung/Blumen</i>	125
5.10	<i>Liste der Variablen-tripel mit der besten accuracy gemittelt über alle Aufgaben - Experiment 2a</i>	127
5.11	<i>Liste der besten Variablenkombinationen mit Angaben zu Accuracy, Precision (Prec) und Recall (Rec) für die einzelnen Wortgenerierungsaufgaben - Experiment 2b</i>	129
5.12	<i>Ergebnisse der Meta-Klassifikation in Experiment 3 - Bei der Verwendung von vier Aufgaben und Majority Voting sind zwei Werte angegeben, da es zwei Stimmen für jede Gruppe gab und somit zwei mögliche Entscheidungen. In der Aufgabe LEX_H/T erzielten zwei Variablenkombinationen die besten Ergebnisse, daher wurden beide untersucht. Kombination 1 beinhaltet die Variablen highest_frequency, shortest_break, avg_google, Kombination 2 die Variablen highest_phon, second_interval, avg_syll.</i>	133
6.1	<i>Beispiel 1 aus der Spontansprache eines Patienten mit chronisch-progredienter Multipler Sklerose; Thema: Hobby; Kontext: Malen bzw. Kunst als eine der wichtigsten Freizeitbeschäftigungen des Patienten</i>	157
6.2	<i>Beispiel 2 aus der Spontansprache eines Patienten mit chronisch-progredienter Multipler Sklerose; Thema: Arbeitsalltag; Kontext: Welche Anstrengungen werden für die Akquise von Finanzmitteln unternommen?</i>	158
6.3	<i>Leistungen des Patienten 01MS_CP in den simplen und komplexen lexikalischen und semantischen Wortgenerierungsaufgaben sowie bei der Verbgenerierung; N. gesamt = Anzahl Nennungen gesamt; N. korrekt = Anzahl Nennungen korrekt; FA(%) = Fehleranteil in Prozent</i>	159
6.4	<i>Beispiel aus der Verbgenerierung eines Patienten mit chronisch-progredienter Multipler Sklerose</i>	159

Abbildungsverzeichnis

2.1	Prozessverlauf - Mannheimer Regulationstheorie; aus Herrmann (2003, S. 225)	9
2.2	Gliederung der Regionen des Frontallappens; aus Thier (2006, S. 471)	12
2.3	Anteile des präfrontalen Kortex (blaue Markierung); aus Kammer & Karnath (2006, S. 491)	13
2.4	Drei Komponenten-Modell des Arbeitsgedächtnisses nach Baddeley und Hinch; aus Baddeley (2003b, S. 830)	14
2.5	Beschreibungsebenen exekutiver Funktionen; aus Drechsler (2007, S. 236)	17
2.6	Sprachregionen im menschlichen Gehirn. 1 Broca-Area; 2 motorische Gesichtsregion; 3 somatosensorische Gesichtsregion; 4 Hörfelder, 5 Wernicke-Area; 6 Gyrus supramarginalis; 7 Gyrus angularis; 8 visuelle Assoziationsregion; rot Äste der A. cerebri media (W. Huber, Aachen); aus Poeck & Hacke (2001, S. 167)	28
2.7	Darstellung der Verlaufsformen bei Multipler Sklerose (modifiziert); aus Pusswald & Vass (2011, S. 332)	35
3.1	Aufgabe LEX_M: Anzahl korrekter Nennungen	66
3.2	Aufgabe SEM_Tiere: Anzahl korrekter Nennungen	66
3.3	Aufgabe LEX_H/T: Anzahl korrekter Nennungen	67
3.4	Aufgabe SEM_Kleidung/Blumen: Anzahl korrekter Nennungen	67
3.5	Verbgenerierung: Anzahl korrekter Nennungen	69
3.6	Aufgabe LEX_M: Anteil korrekter Nennungen, Perseverationen und einfacher Regelbrüche in allen Gruppen	71
3.7	Aufgabe SEM_Tiere: Anteil korrekter Nennungen, Perseverationen und einfacher Regelbrüche in allen Gruppen	72
3.8	Aufgabe LEX_H/T: Anteil korrekter Nennungen, Perseverationen, einfacher Regelbrüche, Kategorieperseverationen und Kategoriefehler in allen Gruppen	72

3.9	Aufgabe SEM_Kleidung/Blumen: Anteil korrekter Nennungen, Perseverationen, einfacher Regelbrüche, Kategorieperseverationen und Kategoriefehler in allen Gruppen	73
3.10	Verbgenerierung: Anteil korrekter Nennungen, Perseverationen und Regelbrüche in allen Gruppen	75
4.1	Workflow: Datenerhebung (<i>data acquisition</i>), Annotation und Aufbereitung der Daten (<i>Annotation/Preprocessing</i>), Extraktion relevanter Variablen (<i>Feature extraction</i>), Variablenauswahl (<i>Feature selection</i>), Training, Klassifikator (<i>Classifier</i>), Entscheidung bzw. Ergebnis der Klassifikation (<i>Patient?</i>); aus Gaspers et al. (2012, S. 211)	87
5.1	Workflow: Datenerhebung (<i>Administration of tests</i>), Transkription/Annotation, Liste der Nennungen (<i>List of responses</i>), Extraktion der Variablen (<i>Feature extraction</i>), Variablenvektoren (<i>Feature vectors</i>), Experimente; aus Thiele et al. (under revision, S. 5)	111
5.2	Meta-Klassifikation mit <i>Majority Voting</i> : SEM_Tiere (<i>Animals</i>), LEX_H/T (<i>Letters H/T</i>), SEM_Kleidung/Blumen (<i>Clothing/Flowers</i>), Trainingsdatensatz (<i>Training Data</i>), Testinstanzen/ -beispiele (<i>Test instances</i>), Training (<i>train</i>), Klassifikatoren (<i>Classifiers</i>), Patienten mit KKD (<i>CCD</i>), Sprachgesund (<i>healthy</i>); aus Thiele et al. (under revision, S. 20) . . .	132
5.3	Vorschlag eines möglichen methodischen Vorgehens: Datenerhebung (<i>Data Collection</i>), Erhebung der Wortgenerierungsdaten (<i>Administer Tests</i>), Annotation der Daten (<i>Annotate data</i>), Evaluation der Variablen (<i>Feature evaluation</i>), Extraktion relevanter Variablen aus den Daten (<i>Extract features</i>), Bestimmung der besten Variablenkombinationen (<i>Determine best combinations</i>), Bestimmung der am besten geeigneten Wortgenerierungsaufgaben (<i>Determine best suited tasks</i>), Trainieren des Meta-Klassifikators (<i>Meta-classifier training</i>), Trainieren der Klassifikatoren (<i>Train classifiers</i>), Zusammenführen der Klassifikatoren (<i>Combine classifiers</i>); aus Thiele et al. (under revision, S. 24);	137

Kapitel 1

Einleitung

”The ability to communicate requires a complex interaction between cognition and language.”

(Coelho & DeRuyter, 1996, S. S5)

Erfolgreiche Kommunikation beruht auf dem Zusammenspiel sprachlicher und kognitiver Funktionen (Coelho & DeRuyter, 1996).¹ Kommunikation kann somit als komplexer Prozess verstanden werden, der, eingebettet in einen situativen Kontext auf der Interaktion sprachlicher, kognitiver und emotionaler Funktionen beruht (Coelho, DeRuyter, Kennedy, & Stein, 2008; Coelho & DeRuyter, 1996). Dieser Prozess kann auf vielfältige Art und Weise gestört sein und zu unterschiedlichen Beeinträchtigungen der kommunikativen Kompetenz mit Auswirkungen auf den beruflichen und gesellschaftlichen Alltag führen (Duff & Proctor, 2002).

Die vorliegende Dissertation widmet sich der Untersuchung sprachlicher sowie kommunikativ-kognitiver Beeinträchtigungen unterschiedlicher Genese und befasst sich dabei insbesondere mit der Analyse von Wortgenerierungsleistungen. Das Ziel ist dabei, eine reliable Identifizierung verschiedener Störungsbilder zu erreichen, um damit eine entsprechende Diagnosestellung zu fazitätieren.

¹Manch einer würde an dieser Stelle darüber hinaus den Stellenwert von Emotionen betonen (Fiehler, 1990; Thiele, Malchus, Jaecks, & Stenneken, 2012; Malchus, Thiele, Jaecks, & Stenneken, 2012), was durchaus seine Berechtigung hat. Eine zusätzliche Auseinandersetzung mit diesem Aspekt würde im Rahmen dieser Arbeit jedoch zu weit führen.

1.1 Motivation & Ziele

Nach einer Hirnschädigung gilt die kommunikative Kompetenz als eine der wichtigsten Prädiktoren für eine erfolgreiche Rückkehr in den Beruf (Wong, Murdoch, & Whelan, 2010). Insbesondere Leistungen in den Exekutivfunktionen werden zu den kognitiven Fähigkeiten gezählt, die eine erfolgreiche berufliche Wiedereingliederung bedingen und die Partizipation am gesellschaftlichen Leben ermöglichen (Lê, Mozeiko, & Coelho, 2011; Ylvisaker & Szekeres, 1989; Duff & Proctor, 2002). Ein Zusammenhang von Störungen in den Exekutivfunktionen und einer beeinträchtigten Kommunikationsfähigkeit liegt somit nahe und wird auch von verschiedenen Autoren berichtet (Douglas, 2010; Martin & McDonald, 2003). Die beobachtbaren Defizite werden häufig als kommunikativ-kognitiv, kognitiv-linguistisch oder pragmatisch bezeichnet (Cummings, 2007; Lê et al., 2011).

Die Diagnosestellung ist bei diesen Defiziten bislang noch eine Hausforderung und nicht immer reliabel möglich. Kommunikativ-kognitive Beeinträchtigungen und andere vergleichsweise subtile Defizite sind häufig nur schwer zu identifizieren und von den Leistungen Sprachgesunder abzugrenzen. Neben einem Mangel an geeigneten Verfahren und Prozeduren zur Diagnose von kommunikativ-kognitiven Defiziten, wird die Identifizierung außerdem durch einen fehlenden Konsens im Hinblick auf die Definition spezifischer Funktionen und beteiligter Prozesse erschwert. So existieren unterschiedliche Terminologien und es fehlt zudem an umschriebenen Definitionen sowohl auf der Ebene spezifischer kognitiver Konstrukte (z.B. Exekutivfunktionen) als auch in Bezug auf die resultierenden Defizite. So werden verschiedene Termini, wie z.B. *high-level-language*-Defizit, subtile (sprachliche) Beeinträchtigung, kognitiv-linguistisches Defizit, kommunikativ-kognitive Beeinträchtigung oder pragmatische Störung oftmals synonym, zum Teil aber auch im Zusammenhang mit spezifischen Ätiologien oder Läsionslokalisationen verwendet. Darüber hinaus fehlt es bislang noch an geeigneten und umfassenden Erklärungsmodellen, welche die aus der Interaktion von Sprache, Kognition und Emotion resultierenden Defizite angemessen abbilden können.

Die Auseinandersetzung mit diesen diagnostischen Herausforderungen bei der Identifizierung verschiedener Sprach- und Kommunikationsstörungen ist insgesamt von großer Relevanz, da eine reliable Diagnose und umfassende Charakterisierung verschiedener Störungsbilder den Ausgangspunkt für die Planung des weiteren Rehabilitationsverlaufes darstellt und für die Auswahl der therapeutischen Interventionen unabdingbar ist (Tesak, 2006; Wehmeyer & Grötzbach, 2006). Da die Prävalenz der Personen mit Hirnschädigung insgesamt steigt (DGN, 2009) und auch subtile Beeinträchtigungen die Teilhabe am gesellschaftlichen Leben beeinflussen und die Wiedereingliederung in den Beruf behindern können (Duff & Proctor, 2002), ergibt sich daraus

auch eine hohe sozio-ökonomische Relevanz für das öffentliche Gesundheitssystem.

Um die Herausforderung einer reliablen Diagnosestellung zu meistern, ist es wichtig, sich auf die Mechanismen, Prozesse oder Funktionen zu fokussieren, die einer spezifischen Beeinträchtigung zugrunde liegen. Dazu braucht eine gute Diagnostik reliable und interpretierbare Maße sprachlicher sowie kognitiver Funktionen (Keil & Kaszniak, 2002). Die Evaluation von Leistungen bei der Wortgenerierung stellt insofern eine interessante Methodologie dar, als eine adäquate Lösung von Wortgenerierungsaufgaben sowohl sprachsystematische als auch kognitiv-exekutive Kompetenzen erfordert (Pekkala, 2012).

Im Rahmen dieser Dissertation wird die Evaluation von Leistungen in unterschiedlichen Wortgenerierungsaufgaben zur Diagnose von sprachlichen und kommunikativ-kognitiven Defiziten unterschiedlicher Genese herangezogen. Neben traditionellen Methoden der Evaluation von Wortgenerierungsleistungen werden etablierte Techniken aus dem Bereich der Informatik (d.h. *machine learning*) zur Evaluation der Wortgenerierungskompetenz herangezogen. Das übergeordnete Ziel ist dabei, einer reliablen Identifizierung und Diagnose kommunikativ-kognitiver Defizite unterschiedlicher Genese durch die Evaluation quantitativer sowie qualitativer Charakteristika in verschiedenen Aufgaben zur Erfassung der Wortgenerierungsfähigkeit näher zu kommen.

Im folgenden Abschnitt werden Aufbau und Inhalt der vorliegenden Arbeit dargelegt.

1.2 Aufbau & Inhalt

Die vorliegende Arbeit gliedert sich in insgesamt sechs Kapitel. Diesem einleitenden Kapitel folgt im nächsten die Darstellung des theoretischen Gerüsts dieser Dissertation. Neben einer Heranführung an theoretische Vorstellungen zum Zusammenspiel sprachlicher und kognitiver Funktionen beinhaltet Kapitel 2 eine fundierte Darstellung des Konstrukts der Exekutivfunktionen einschließlich einer Definition sowie Erläuterungen zu neuronalen Korrelaten und modelltheoretischen Aspekten. Daran anschließend folgt eine theoretische Abhandlung zur Wortgenerierungsfähigkeit mit einer Darstellung der typischerweise zur Elizitierung verwendeten Wortgenerierungsaufgaben sowie der kognitiven Funktionen die ihnen zugrunde liegen. Im Weiteren werden relevante Grundlagen zu den im Rahmen dieser Arbeit untersuchten sprachsystematischen und kommunikativ-kognitiven Beeinträchtigungen sowie der Multiplen Sklerose und mit ihr in Zusammenhang stehenden Defi-

ziten in Sprache, Kommunikation und Kognition geliefert. Darauf folgt eine Beschreibung wichtiger Aspekte der Diagnostik von Sprach- und Kommunikationsstörungen sowie eine Heranführung an den Einsatz von Techniken aus dem maschinellen Lernen² zur Identifizierung spezifischer Störungen. Den Abschluß bildet eine kurze Darstellung der studienübergreifenden Fragestellungen.

In Kapitel 3, 4 und 5 werden daran anschließend drei empirische Studien vorgestellt, die sich mit dem Einsatz verschiedener Methodologien zur Evaluation von Wortgenerierungsleistungen in verschiedenen klinischen Populationen befassen. Die erste Studie (siehe Kapitel 3) befasst sich mit der Evaluation quantitativer sowie qualitativer Charakteristika der Wortgenerierung bei Patienten mit verschiedenen Verlaufsformen einer Multiplen Sklerose im Vergleich zu Sprachgesunden. Die zweite Studie (siehe Kapitel 4) beinhaltet eine Evaluation von Wortgenerierungsleistungen zur Klassifikation von Sprach- und Kommunikationsstörungen mit *machine learning*-Techniken. Neben traditionellen Evaluationsmaßen werden hier weitere lexikalische, sublexikalische und zeitliche Variablen bei der Evaluation der Wortgenerierungsleistungen berücksichtigt. Mit der dritten Studie (siehe Kapitel 5) wird die Entwicklung einer Methodologie zur Evaluation von Wortgenerierungsleistungen mit *machine learning*-Techniken exploriert. Die Studie befasst sich mit der Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden und stellt eine Erweiterung der Analysen aus der zweiten Studie dar.

Im Anschluss an die Beschreibung der drei Studien erfolgt in Kapitel 6 eine studienübergreifende Diskussion. Diese beginnt mit einer kurzen Zusammenfassung der wichtigsten Ergebnisse, gefolgt von einer Diskussion der Relevanz quantitativer und qualitativer Auswertungsparameter, sowie der Evaluation von Leistungen in verschiedenen Wortgenerierungsaufgaben. Daran anschließend folgt die Darstellung eines Einzelfalls, der sozial-kognitive Aspekte bei der Evaluation von Wortgenerierungsleistungen beleuchtet. Abschließend soll der Einsatz von *machine learning*-Techniken bei der Evaluation von Wortgenerierungsleistungen diskutiert werden.

Auf diese studienübergreifende Diskussion aufbauend folgen dann ein abschließendes Fazit bezüglich der Konsequenzen für das diagnostische Vorgehen sowie ein Ausblick auf weitere Forschung zur Evaluation von Wortgenerierungsleistungen.

²engl.: *machine learning*

Kapitel 2

Wortgenerierung im klinisch-linguistischen Kontext

Die Fähigkeit zu kommunizieren beruht nicht allein auf einem funktionierenden Sprachsystem: sprachliche Äußerungen werden nicht in einem linguistischen Vakuum produziert, sondern eingebettet in eine spezifische Situation, einen kommunikativen Kontext und, betrachtet man Kommunikation vorrangig als dialogischen Prozess, in Interaktion mit einem (oder mehreren) Kommunikationspartner(n) (Cummings, 2007). Erfolgreiche Kommunikation erfordert daher neben der korrekten Anwendung sprachlicher Regeln, das heißt neben rein sprachsystematischen Fähigkeiten, auch die Integration von Wissen über den Gesprächspartner, den kommunikativen Kontext sowie allgemeines Weltwissen (Martin & McDonald, 2003; Douglas, 2010). Dieses ist beispielsweise erforderlich um in der sozial-kommunikativen Interaktion Inferenzen ziehen und Äußerungen hinsichtlich ihrer mitunter non-literalen Bedeutung korrekt enkodieren zu können. Man spricht in diesem Zusammenhang auch von pragmatischen Kompetenzen (Douglas, 2010; Lê et al., 2011; Martin & McDonald, 2003), welche Turkstra, McDonald und Kaufmann (1995) folgendermaßen definieren:

”Pragmatic communication skills are the skills underlying competence in contextually determined, functional language use.”
(Turkstra et al., 1995, S. 329)

Pragmatische Kompetenzen bezeichnen demnach die Fähigkeit zur Anwendung von Sprache in einem kommunikativen Kontext. Sprache wird dabei als eine Möglichkeit des zielorientierten Handelns beteiligter Personen in alltäglichen Kommunikationssituationen verstanden (Cummings, 2007). Dabei betonen verschiedene Autoren die Relevanz der Interaktion sprachlicher

und nicht-sprachlicher kognitiver Funktionen für eine erfolgreiche Kommunikation (Douglas, 2010; Martin & McDonald, 2003; Perkins, 2005; Turkstra et al., 1995). So beschreibt Perkins (2005) pragmatische Fähigkeiten beispielsweise als Konsequenz der Interaktion von linguistischen, kognitiven und sensomotorischen Funktionen, welche sowohl in einem Individuum als auch zwischen Individuen, das heißt intra- und interindividuell, stattfindet. Pragmatische Fähigkeiten dienen dazu Äußerungen unter Berücksichtigung des situativen Kontextes an die Bedürfnisse des jeweiligen Gesprächspartners (z.B. abhängig von dessen Vorwissen) anzupassen und sie in einer logischen und kohärenten Art und Weise zu verbalisieren. Dazu zählt auch, die eigenen Äußerungen bezüglich ihrer Angemessenheit zu kontrollieren und insgesamt Verantwortung für ein Gelingen der Kommunikation und das Erreichen des damit einhergehenden (Handlungs-)Ziels zu übernehmen (Turkstra et al., 1995). Darüber hinaus existieren weitere Ansätze zur Definition pragmatischer Fähigkeiten. So werden beispielsweise in einem eher modularen Ansatz pragmatische Fähigkeiten als eine Komponente im Sprachsystem, ähnlich der Syntax oder der Semantik, beschrieben (vgl. Joannette & Ansaldo, 1999). Andere wiederum beschreiben pragmatische Fähigkeiten als ein rein kognitives Modul (Wilson, 2005). Die Verschiedenheit dieser Annahmen verdeutlicht die Kontroversen, die noch bezüglich einer einheitlichen Definition pragmatischer Fähigkeiten bestehen. Im Rahmen dieser Dissertation wird (in Anlehnung an Perkins, 2005) ein interaktionistischer Ansatz pragmatischer Fähigkeiten vertreten, der von einem engen Zusammenspiel kognitiver und linguistischer Prozesse ausgeht.

Illustriert werden kann dieses Zusammenspiel anhand von Evidenzen aus der klinischen Pragmatik (Cummings, 2007). Entsprechende Defizite werden bei klinischen Populationen diverser Ätiologien wie z.B. Autismus, Schizophrenie, Schädelhirntrauma, rechts-hemisphärischen oder linkshemisphärischen Läsionen, Demenz etc. beschrieben (siehe Cummings, 2007; vgl. auch Abschnitt 2.4.2) und häufig als pragmatische, kognitiv-linguistische oder kommunikativ-kognitive Störung bezeichnet³ (Lê et al., 2011; Turkstra et al., 1995). Als häufige Symptome werden unter anderem ausschweifende Äußerungen, häufige Themenwechsel, ungewöhnliche Wortwahl, Mangel an Kohärenz, Schwierigkeiten bei der Disambiguierung mehrdeutiger Äußerungen sowie beim Schlussfolgern und Verstehen non-literaler Bedeutungen genannt, während sprachstrukturelle Funktionen meist erhalten sind (vgl. Cummings, 2007; Douglas, 2010; Lê et al., 2011; Martin & McDonald, 2003). Kommunikativ-kognitive Störungen werden unter anderem im Zusammenhang mit Defiziten in den Exekutivfunktionen beschrieben (Dou-

³Die Bezeichnung *kommunikativ-kognitiv* wird in dieser Arbeit vorrangig verwendet. Dieser Begriff wurde gewählt um kenntlich zu machen, dass sich die damit bezeichneten Defizite in der Kommunikation manifestieren jedoch eine kognitive Genese haben.

glas, 2010; Martin & McDonald, 2003). So konnten Douglas und Kollegen (2010) in einer Studie mit Schädelhirntrauma-Patienten einen Zusammenhang zwischen pragmatischen Beeinträchtigungen in der Kommunikation und Defiziten in den Exekutivfunktionen zeigen. Die Relevanz einer reliablen Identifizierung und adäquaten Behandlung kommunikativ-kognitiver Beeinträchtigungen ergibt sich insbesondere daraus, dass diese Störungen zu Einschränkungen im Alltag sowie auch im Berufsleben führen und somit einen erheblichen Einfluss auf die Lebensqualität betroffener Patienten haben.

2.1 Modelltheoretische Perspektiven

Insbesondere vor dem Hintergrund einer Weiterentwicklung diagnostischer Möglichkeiten und daraus resultierender therapeutischer Interventionen ist eine modelltheoretische Einbettung von Symptomen und spezifischen Störungsbildern sinnvoll. Nach Uomoto (1991) ist es wichtig, die im Rahmen einer Diagnostik identifizierten Symptome auf Basis eines Modells kognitiver Funktionen zu interpretieren. Diese Forderung lässt sich auch auf die Evaluation von sprachlichen oder kommunikativ-kognitiven Leistungen übertragen.

So wichtig wie diese Forderung ist, so schwierig ist sie auch zu erfüllen. Während für die Einbettung sprachsystematischer Störungen auf Wort- oder Satzebene verschiedene Modelle herangezogen werden können (z.B. das Logogen-Modell nach Morton (1969) oder Ellis (1982); das Sprachverarbeitungsmodell nach Levelt (1998); konnektionistische Modelle der Sprachverarbeitung nach Schade (1992) oder Dell (1999); das Satzproduktionsmodell nach Garrett (1984) etc.) und zum Teil sogar modellspezifische Diagnostiken existieren (z.B. Lexikon Modellorientiert (LeMo), De Bleser, 2004, auf Basis des Logogenmodells), gibt es nur wenige und häufig noch recht unspezifische Modelle, die sich mit Störungen an der Schnittstelle sprachlicher und kognitiver Funktionen beschäftigen. Beispielhaft sind hier zu nennen: das *3-Ebenen-Modell der sprachlichen Verarbeitung* in Huber & Ziegler (2000), das *Model of Brain Behavior Relationship* von Reitan & Wolfson (1988) oder das hierarchische *Modell höherer kortikaler Funktionen* von Stuss & Benson (in Uomoto, 1991).

Eine eher sprachpsychologische Betrachtung der Verarbeitung von Sprache, welche verbale Äußerungen eines Individuums in einer spezifischen Kommunikationssituation fokussiert, findet sich in der Mannheimer Regulationstheorie (Herrmann & Grabowski, 1994; Herrmann, 2003). Den Autoren gelingt damit eine Verknüpfung sprachbezogener und kognitiver Aspekte (Herrmann, 2003). Die im Folgenden dargestellte Theorie soll die Relevanz eines gemeinsamen Modells für Sprache und Kognition hervorheben.

2.1.1 Mannheimer Regulationstheorie der Sprachverarbeitung (Herrmann & Grabowski, 1994)

Die Wichtigkeit der Kognition bei der Sprachverarbeitung stellen Herrmann und Grabowski in der Mannheimer Regulationstheorie heraus (Herrmann & Grabowski, 1994)⁴. Die Autoren sehen Sprachverarbeitung als integriert in kognitive bzw. psychische Prozesse und beschreiben mehr oder weniger eine wechselseitige Beziehung zwischen Sprache und Kognition. Wesentlich für diese Theorie ist die Annahme, dass die Produktion sprachlicher Äußerungen immer im Hinblick auf ein bestimmtes Ziel erfolgt. Wie jedes andere Verhalten auch wird eine Äußerung dann produziert, wenn mit ihr ein bestimmtes Ziel erreicht werden soll (Rummer, 1996). Dabei wird betont, dass die Sprachproduktion nicht nur auf der Anwendung sprachstruktureller Regeln beruht sondern in besonderen Maße auch von der Integration kontextueller Information abhängt, die, zum Beispiel bei der Produktion von Redewendungen oder Ellipsen, eine Äußerung verstehbar macht (Herrmann, 2003).

Sprachliche Äußerungen sind nach Herrmann und Grabowski (Herrmann & Grabowski, 1994) das Ergebnis eines Regulationsprozesses. "Sprechen ist dabei ein Mittel zur Angleichung von Ist- und Sollzuständen im kognitiven *System* des Sprechers" (Rummer, 1996, S. 6). Ist die Diskrepanz zwischen Ist- und Sollzuständen zu groß wird bei einem Teil der Ist-Soll-Abweichungen⁵, eine Stelloperation angetrieben, die in einer sprachlichen Äußerung resultiert. Bei den sprachproduktionsrelevanten Ist-Soll-Abweichungen werden zwei Typen unterschieden, die Stelloperationen erforderlich machen. Zu diesen gehören (1) Handlungsziele, die noch nicht erreicht sind und am wahrscheinlichsten mit der Produktion einer sprachlichen Äußerung erreicht werden können, und (2) Situationen, in denen bestehende Konventionen eine sprachliche Äußerung erfordern (Herrmann, 2003). Die Regulation gilt als erfolgreich, wenn durch die sprachliche Äußerung das Handlungsziel unter Einhaltung bestehender Konventionen erreicht wird (Herrmann, 2003). Der Prozess der Sprachproduktion ist in Abhängigkeit vom situativen Kontext somit sehr variabel; die Kommunikationssituation bestimmt mehr oder weniger den Prozess der Sprachproduktion. Es wird deutlich, dass sich diese Theorie der Sprachverarbeitung auf das in einer Situation agierende Individuum fokussiert und nicht, wie viele andere Modelle, auf eher sprachstrukturelle Aspekte wie etwa die morphologische Enkodierung oder den Wortformabruf aus dem Lexikon (Rummer, 1996).

⁴Die Mannheimer Regulationstheorie (Herrmann & Grabowski, 1994) soll im Rahmen dieser Arbeit nicht zur Einbettung komplexer sprachlicher und/oder kommunikativ-kognitiver Defizite herangezogen oder gar erweitert werden. Sie dient lediglich als Beispiel einer Theorie, welche sprachliche und kognitive Aspekte der Sprachverarbeitung modelltheoretisch berücksichtigt.

⁵Andere Formen von Ist-Soll-Abweichungen können Stelloperationen hervorrufen, die in nicht-sprachlichen Handlungen resultieren (Herrmann, 2003).

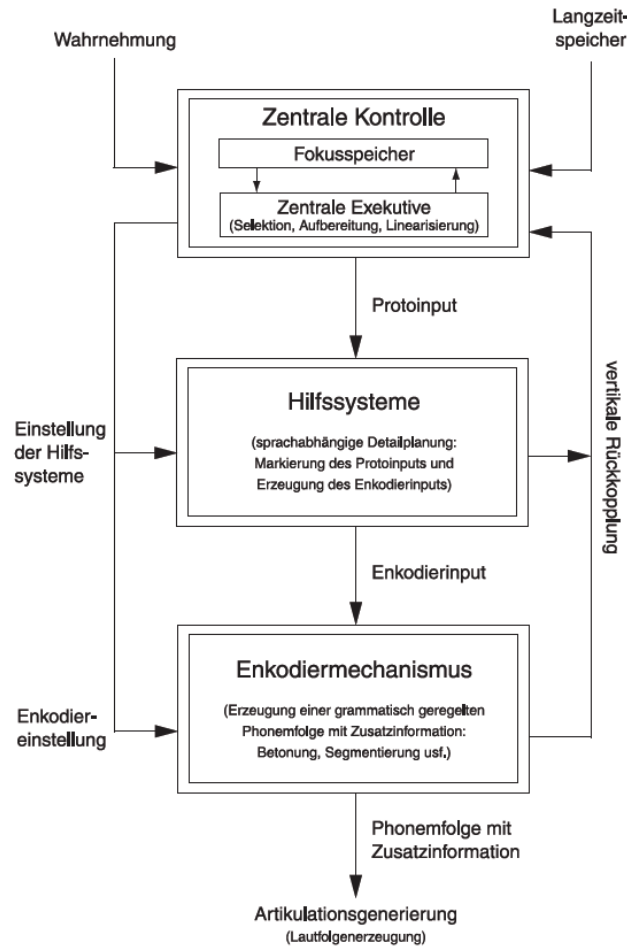


Abbildung 2.1: Prozessverlauf - Mannheimer Regulationstheorie; aus Herrmann (2003, S. 225)

In der Mannheimer Regulationstheorie werden drei Verarbeitungsebenen unterschieden⁶. Diese sind hierarchisch geordnet und verfügen über ein Rückkopplungssystem von unteren zu höheren Ebenen. Die wichtigsten Instanzen sind dabei die *Zentrale Kontrolle*, die *Hilfssysteme* und der *Enkodiermechanismus*. Abbildung 2.1 stellt die Prozesskomponenten der Mannheimer Regulationstheorie schematisch dar. Während die obere Ebene (*Zentrale Kontrolle*) aufmerksamkeitsgesteuerte, kontrolliert arbeitende Prozesse umfasst, werden die untergeordneten Ebenen (*Hilfssysteme* und *Enkodiermechanismus*) demgegenüber als automatisierte, auf Routinen beruhende Regulationsebenen verstanden (Rummer, 1996).

⁶Es existiert auch eine vierte Ebene, welche phonetisch-artikulatorische Aspekte beinhaltet. Diese wird jedoch nicht näher beschrieben (Herrmann, 2003).

Die *Zentrale Kontrolle* besteht aus zwei Komponenten. Die erste ist ein Arbeitsspeicher (*Fokusspeicher*) mit begrenzter Kapazität, der die für die Produktion verbaler Äußerungen relevanten Informationen über die Zustände des Systems enthält. Anhand der Informationen aus dem *Fokusspeicher* erfolgt die Entscheidung, ob eine Äußerung produziert wird oder nicht (Herrmann, 2003). Die zweite Komponente ist die *Zentrale Exekutive*. Sie ist für die Selektion, Aufbereitung und Linearisierung der Information aus dem *Fokusspeicher* zuständig und übernimmt die prozedurale Organisation. Unter Berücksichtigung des bisherigen Gesprächsverlaufs erstellt die *Zentrale Exekutive* einen für die weitere sprachliche Enkodierung vorgesehenen *Protoinput*, der über eine Propositionsstruktur verfügt (Herrmann, 2003; Rummer, 1996). Auf der Ebene der *Hilfesysteme* wird der vorsprachliche *Protoinput* für die sprachliche Enkodierung weiter vorbereitet. Dieser Zwischenschritt ist erforderlich, da angenommen wird, dass selbst einfache Propositionen von dem relativ unflexiblen *Enkodiermechanismus* nicht automatisch in eine situationsangemessene und sprachstrukturell korrekte Äußerung übersetzt werden können. Die *Hilfesysteme* dienen also der sprachlichen Detailplanung (Herrmann, 2003). So geben verschiedene, parallel arbeitende *Hilfesysteme* unter anderem Informationen über Satzart, Tempus und Modus oder ermöglichen eine situationsangemessene Kohärenz der Äußerung, woraus der *Enkodierinput* hervorgeht. Dieser wird dann auf der Ebene des *Enkodiermechanismus* in eine entsprechende Phonemabfolge übersetzt (Herrmann, 2003). Der *Enkodiermechanismus* arbeitet automatisch und folgt weitestgehend den Prinzipien lokal-konnektionistischer Modelle (Herrmann & Grabowski, 1994; Herrmann, 2003). Output für die Artikulation ist eine Phonemfolge, die darüber hinaus zusätzliche Informationen, beispielsweise bezüglich der Prosodie, enthält. Eine detaillierte Beschreibung der einzelnen Instanzen findet sich z.B. bei Herrmann und Grabowski (1994), Herrmann (2003) oder Rummer (1996).

Die *Zentrale Exekutive* nimmt in der Theorie von Herrmann und Grabowski (1994) eine wichtige Rolle bei der Integration von Wissen aus dem Arbeitsspeicher ein; ihr obliegt die operative Kontrolle des Sprachproduktionsprozesses. Die Annahme einer *Zentralen Exekutive* entspricht dem, was landläufig unter Exekutivfunktionen verstanden wird. Die Schwierigkeit bei der Auseinandersetzung mit Exekutivfunktionen sowie mit an der Schnittstelle von sprachlichen und kognitiven Funktionen liegenden Defiziten besteht darin, dass zum Teil noch keine einheitliche Terminologie existiert und Kontroversen bezüglich Definition und Abgrenzung bestehen.

Der nun folgenden Abschnitt widmet sich einer eingehenden Darstellung des Konstrukts der Exekutivfunktionen und berücksichtigt dabei neben definitionsbezogenen Aspekten und neuroanatomischen Korrelaten kognitiv-exekutiver Funktionen auch bekannte modelltheoretische Ansätze.

2.2 Exekutivfunktionen

Exekutivfunktionen gelten als ein psychologisches Konstrukt, welches unter anderem als relevant für die Kommunikation bzw. den kommunikativen Erfolg beschrieben wird (Coelho & DeRuyter, 1996; Drechsler, 2007; Lê et al., 2011; Ylvisaker & Szekeres, 1989). Lange Zeit bestanden Kontroversen bezüglich der Frage, ob den Exekutivfunktionen ein basaler Mechanismus zugrunde liegt; heutzutage nehmen die meisten Autoren jedoch unabhängige Komponenten von Exekutivfunktionen an (Alvarez & Emory, 2006; Drechsler, 2007; Fuster, 2002; Keil & Kaszniak, 2002; Miyake, 2000; Stuss & Alexander, 2000, 2007; Stuss, 2011). Im Folgenden sollen theoretische Annahmen und Modelle sowie neuroanatomische Aspekte von Exekutivfunktionen vorgestellt werden. Dabei wird deutlich werden, dass das Konstrukt der Exekutivfunktionen noch nicht im Detail entschlüsselt ist und noch immer Kontroversen bezüglich relevanter Komponenten bestehen. Es lassen sich jedoch auch Gemeinsamkeiten in den (modell-)theoretischen Annahmen erkennen, die hier herausgearbeitet werden sollen.

2.2.1 Definition

Unter Exekutivfunktionen werden im Wesentlichen Prozesse höherer Ordnung verstanden, die für Kontrolle, Regulation und (zeitliche) Integration von spezifischen kognitiven Prozessen zuständig sind, um zielgerichtetes und situationsangepasstes Handeln zu ermöglichen (Alvarez & Emory, 2006; Drechsler, 2007; Fuster, 2002; Miyake, 2000). Exekutivfunktionen sind zuständig für die Koordination von motiviertem und angepasstem Verhalten und ermöglichen einen flexiblen Einsatz kognitiver Fähigkeiten, um so neue oder anspruchsvolle Aufgaben meistern zu können (Martin & McDonald, 2003). Eine endgültige Definition von Exekutivfunktionen hat sich bis heute nicht durchgesetzt. Vielmehr sind, basierend auf neuropsychologischen und anatomischen Theorien, Aufzählungen und Beschreibungen von Komponenten, Mechanismen oder Basisprozessen zu finden, welche als relevante Bestandteile exekutiver Funktionen diskutiert werden (Baddeley, 2003b; Repovš & Baddeley, 2006; Fuster, 2002; Miyake, 2000; Shallice, 2002; Smith & Jonides, 1999). Einige der im Zusammenhang mit Exekutivfunktionen diskutierten Theorien sollen im Folgenden vorgestellt werden.

2.2.2 Neuroanatomische Korrelate

Insbesondere das Frontalhirn, das heißt der frontale bzw. präfrontale Kortex (siehe Abbildung 2.2), wird mit Exekutivfunktionen in Verbindung gebracht.

Exekutivfunktionen werden in der Literatur häufig auch als Frontalhirnfunktionen bezeichnet (Alvarez & Emory, 2006; Drechsler, 2007). Diese Ver-

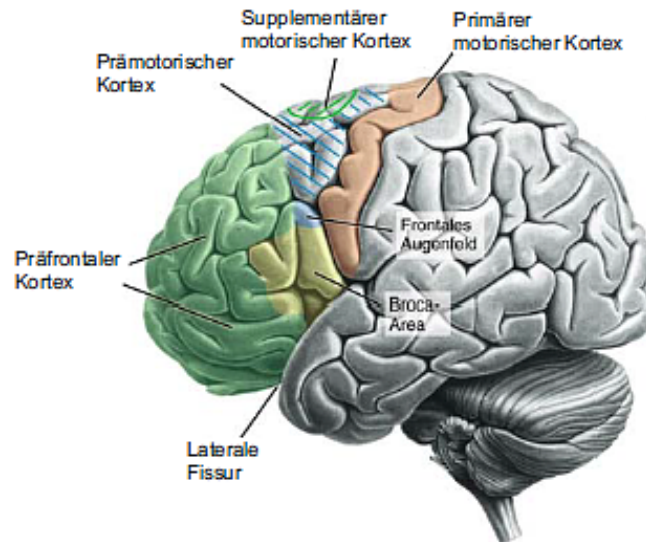


Abbildung 2.2: Gliederung der Regionen des Frontallappens; aus Thier (2006, S. 471)

mischung funktioneller und anatomischer Beschreibungen ist jedoch insofern schwierig, als Frontalhirnschädigungen nicht zwingend zu Störungen der Exekutivfunktionen führen und auch diffuse, nicht-frontale Läsionen Defizite in den Exekutivfunktionen hervorrufen können (Alvarez & Emory, 2006; Keil & Kaszniak, 2002; Stuss & Alexander, 2000). So werden Beeinträchtigungen in den Exekutivfunktionen beispielsweise auch nach Schädigung der Basalganglien, des Thalamus oder bei diffuser axonaler Schädigung der weißen Substanz berichtet (Heidler, 2009; Keil & Kaszniak, 2002; Wong et al., 2010). Es wird angenommen, dass es sich bei Exekutivfunktionen nicht um eine einzeln anatomisch lokalisierte Funktion handelt, sondern, dass verschiedene, in unterschiedlichen Hirnarealen lokalisierte Prozesskomponenten zusammenwirken (Keil & Kaszniak, 2002). Dieses wird auch durch die komplexe Verschaltung des frontalen Kortex mit anderen Hirnstrukturen deutlich.

Der frontale Kortex besitzt zahlreiche Verbindungen zu kortikalen und subkortikalen Strukturen. Nach Fuster (2002) handelt es sich beim präfrontalen Kortex um die am besten verknüpfte Struktur des Kortex überhaupt. Die spezifische Funktion einzelner Verbindungen des präfrontalen Kortex ist nicht gänzlich bekannt, wird aber häufig ausgehend von den mit dem präfrontalen Kortex verbundenen Strukturen inferiert (Fuster, 2002).

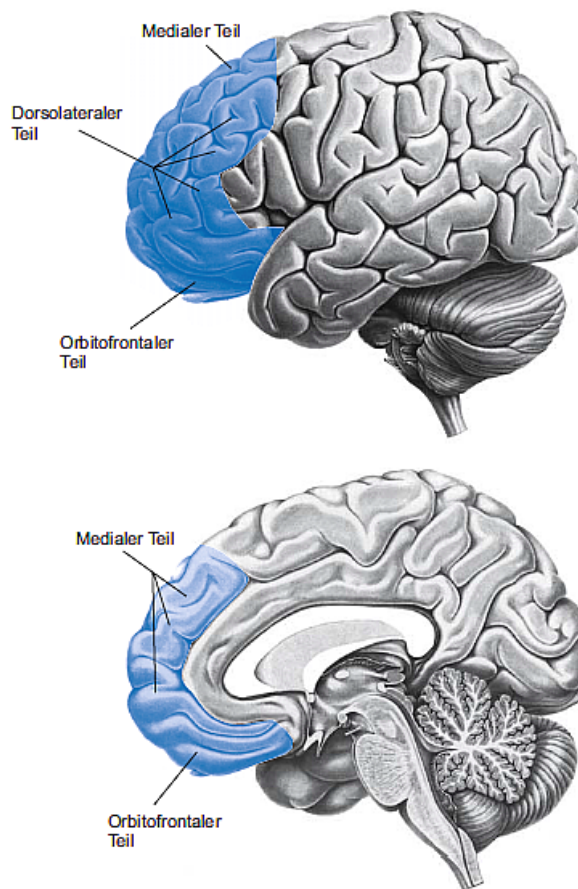


Abbildung 2.3: Anteile des präfrontalen Kortex (blaue Markierung); aus Kammer & Karnath (2006, S. 491)

Mediale, laterale und orbitale Anteile des frontalen Kortex (siehe Abbildung 2.3) sind reziprok miteinander sowie mit Kernen des anterioren und dorsalen Thalamus verbunden (Fuster, 2002). Alvarez und Emory (2006) unterscheiden drei verschiedene fronto-subkortikale Schaltkreise: einen dorsolateralen, einen ventromedialen sowie einen orbitofrontalen. Von diesen wird insbesondere der dorsolaterale Schaltkreis in Zusammenhang mit Exekutivfunktionen beschrieben. Die Projektionen vom dorsolateralen frontalen Kortex führen zum dorsalen Kopf des Nucleus Caudatus und sollen unter anderem bei der verbalen Flüssigkeit, beim Planen, Problemlösen, bei der Inhibition sowie beim Arbeitsgedächtnis, beim abstrakten Denken und beim Schlussfolgern eine wichtige Rolle spielen (Alvarez & Emory, 2006). Projektionen des ventromedialen Kortex (zum anterioren Cingulum und dem Nucleus Accumbens) werden mit Motivation in Verbindung gebracht, Projektionen des orbitofrontalen Kortex (zum ventromedialen Anteil des Nucleus Caudatus) mit sozial korrektem Verhalten. Zusätzlich zu diesen di-

rekten Verbindungen verfügt der frontale Kortex über zahlreiche Verbindungen mit anderen kortikalen und subkortikalen Arealen sowie mit Strukturen des Hirnstamms.

Auch nach Duffy und Campbell (2001, in Alvarez & Emory, 2006) ist der frontale Kortex nur ein Teil eines größeren Systems von Exekutivfunktionen, welches auf der Interaktion verschiedener Strukturen des zentralen Nervensystems beruht. Beeinträchtigungen der Exekutivfunktionen sind somit nicht nur nach Schädigung des Frontalhirns selbst zu erwarten sondern auch bei Läsionen im Bereich des weit verzweigten frontal-kortikal-subkortikalen Netzwerks. Das heißt, kognitiv-exekutive Defizite können auch bei Unterbrechung der Verbindung zwischen beteiligten Strukturen oder nach Schädigung der mit dem frontalen Kortex verbundenen Strukturen, beispielsweise nach Basalganglien-, Thalamus- oder Hirnstammläsionen, auftreten.

2.2.3 Modelle & Taxonomie

Exekutivfunktionen werden häufig als Teil von Arbeitsgedächtnisprozessen beschrieben, wie zum Beispiel im Arbeitsgedächtnismodell von Baddeley (2003a; 2003b).

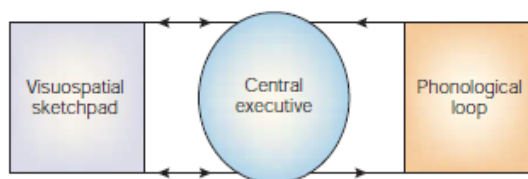


Abbildung 2.4: Drei-Komponenten-Modell des Arbeitsgedächtnisses nach Baddeley und Hinch; aus Baddeley (2003b, S. 830)

Neben einer *zentralen Exekutive*, die als Kontrollinstanz fungiert, werden im ursprünglichen Modell (vgl. Abbildung 2.4) ein sprachliches (*phonologische Schleife*, engl.: *phonological loop*) sowie ein visuell-räumliches (*visuell-räumlicher Skizzenblock*, engl.: *visuospatial sketchpad*) Speichersystem angenommen. In einer Weiterentwicklung des Modells wird eine zusätzliche Modellkomponente, der episodische Speicher (*episodic buffer*) eingeführt, dessen Aufgabe domänenübergreifend in der Integration und Sequenzierung verbaler und visuell-räumlicher Informationen besteht (Baddeley, 2000). Die *zentrale Exekutive* ist für die aufmerksamkeitsgesteuerte Kontrolle des Arbeitsgedächtnisses bzw. der untergeordneten Speichersysteme zuständig und umfasst vier Basiskomponenten: Aufmerksamkeitsfokussierung, Aufmerksamkeitsteilung, Wechsel des Aufmerksamkeitsfokus und Verbindung von Arbeitsgedächtnisinhalten mit dem Langzeitgedächtnis (Repovš & Baddeley, 2006). Die *zentrale Exekutive* ist somit auch in Baddeleys Modell kein

Homunkulus: Exekutivfunktionen werden auch hier als ein aus verschiedenen Komponenten bestehendes Konstrukt verstanden (Repovš & Baddeley, 2006).

Einen weiteren Ansatz liefern Norman und Shallice (1980) bzw. in einer Erweiterung Shallice, Burgess und Robertson (1996) mit dem *Supervisory Attentional System*-Modell (auch: *SAS*-Modell). Die Autoren gehen davon aus, dass es Handlungen gibt, die automatisch (d.h. *routine*) auf Basis von Schemata ablaufen und solche, die nicht-automatisch (d.h. *non-routine*) ablaufen und einer bewussten Kontrolle unterliegen. Das Modell unterscheidet zwei Strukturen: das *Supervisory Attentional System* und den *Contention Scheduler*. Während der *Contention Scheduler* als Prozessor für automatisierte Handlungen, also Routinehandlungen, zuständig ist, stellt das *Supervisory Attentional System* eine Kontrollinstanz dar. Es überwacht die routinierten Abläufe und kommt dann zum Tragen, wenn von Routinehandlungen abgewichen werden muss, eine bewusste Verarbeitung gefordert und eine flexible Anpassung an die Situation erforderlich ist. Das *Supervisory Attentional System* ist demnach in neuen Situationen oder solchen mit wechselnden Anforderungen gefordert, in denen Routinehandlungen nicht mehr ausreichen, um ein Ziel zu erreichen, sondern abweichend davon die Generierung und Anwendung von Strategien sowie Monitoring erforderlich ist (Shallice, 2002). Shallice und Kollegen (1996) beschreiben das *Supervisory Attentional System* als im präfrontalen Kortex lokalisiert, während der *Contention Scheduler* als Teil des striatal-thalamischen Systems gesehen wird. Schädigungen des präfrontalen Kortex führen demnach zu Verhalten, das ausschließlich auf Routineabläufen des *Contention Scheduler* basiert (Shallice, 2002). Insbesondere das Auftreten von Perseverationen, ein Haftenbleiben an Konzepten oder auch inkohärentes Verhalten wird mit einer Störung des *Supervisory Attentional System* in Verbindung gebracht (Müller & Münte, 2000; Shallice, 2002).

Smith und Jonides (1999) beschreiben Exekutivfunktionen als Regulationsprozesse, die mit den Inhalten des Arbeitsgedächtnisses operieren und an komplexen Integrationsprozessen beteiligt sind. Die Autoren nennen fünf Prozesse als Komponenten von Exekutivfunktionen: (i) richten des Aufmerksamkeitsfokus auf relevante und Inhibition irrelevanter Verarbeitungsprozesse (*attention and inhibition*), (ii) zeitliche Integration von Prozessen bei komplexen Aufgaben inklusive Wechseln des Aufmerksamkeitsfokus (*task management*), (iii) zielgerichtete Planung von Teilschritten (*planning*), (iv) Monitoringprozesse, das heißt Aktualisierung und Überprüfung von Arbeitsgedächtnisinhalten zur Planung folgender Handlungsschritte (*monitoring*), (v) Kodierung von Arbeitsgedächtnisrepräsentationen bezüglich Ort und Zeit des Auftretens (*coding*). Die Auflistung dieser Komponenten deckt sich weitestgehend mit denen aus dem Modell von Baddeley (2006) und zeigt auf,

dass das grundsätzliche Verständnis von Exekutivfunktionen verschiedener Autoren ähnlich ist.

Auch Fuster (2002) hebt die Bedeutung der zeitlichen Intergration als Komponente von Exekutivfunktionen hervor. Er sieht die zeitliche Organisation von zielgerichteten Handlungen in Verhalten, Kognition und Sprache als die bedeutendste Exekutivfunktion. Neuroanatomisch lokalisiert ist der Prozess der zeitlichen Integration nach Fuster (2002) im lateralen präfrontalen Kortex. Nach Ansicht des Autors besteht die zeitliche Integration in der Umsetzung von diskontinuierlich wahrgenommenen Reizen und neuralem Input in kohärente Handlungsstrukturen und wird dabei von Funktionen des Arbeitsgedächtnisses, des *preparatory set* und der inhibitorischen Kontrolle unterstützt.

Miyake und Kollegen (2000) beschreiben drei basale Komponenten von Exekutivfunktionen. Die Autoren untersuchten *shifting* (d.h. *mental set shifting* – kognitive Flexibilität), *monitoring/updating* (d.h. Überwachung und Aufrechterhaltung von Repräsentationen des Arbeitsgedächtnisses) und *inhibition* (d.h. Inhibition vorherrschender Antworten) und konnten für diese Basiskomponenten zeigen, dass sie, obwohl sie mäßig miteinander korrelieren, klar voneinander separierbare Komponenten darstellen und in unterschiedlichem Maß an der Bewältigung komplexer Aufgaben beteiligt sind.

Darüber hinaus existieren weitere Theorien, Konzepte und Beschreibungen von Exekutivfunktionen (vgl. auch Constantinidou, Wertheimer, Tsanadis, Evans, & Paul, 2012). Beispielhaft sind hier zu nennen⁷ die Ansätze von Busch, McBride, Curtiss und Vanderploeg (2005; *kognitive Flexibilität, mentale Kontrolle und Arbeitsgedächtnis*) oder Callahan, Johnstone und Stonnington (2001; *Initiierung, Terminierung* im Sinne von Impulskontrolle und *Selbstregulation*) sowie das klinische Modell in Sohlberg und Mattee (2001; *Initiierung, Inhibition, Persistenz* durch Aufrechterhaltung von Aufmerksamkeit und Arbeitsgedächtnisleistungen, *Organisation* im Sinne von Planung, Zielidentifizierung und Timing, *Generatives Denken* durch kognitive Flexibilität, Problemlösen, Ideengenerierung und Perspektivübernahme, *Wahrnehmung* im Sinne von Monitoring und Einsicht).

Den bisher genannten Theorien ist gemein, dass sie kognitive Aspekte von Exekutivfunktionen fokussieren. Dabei scheinen insbesondere Prozesse des Arbeitsgedächtnisses eine wichtige Rolle zu spielen. Darüber hinaus existieren Theorien, die Exekutivfunktionen nicht ausschließlich auf kognitive Funktionen beschränken, sondern beispielsweise auch emotionale Regulationsprozesse sowie Antrieb und Motivation einbeziehen. Mitunter wird daher in der Literatur eine Differenzierung von kognitiven ("cold") und mo-

⁷Die folgende Aufzählung enthält in Klammern jeweils die von den jeweiligen Autoren beschriebenen Komponenten von Exekutivfunktionen.

tivationalen ("hot") Exekutivfunktionen vorgenommen (Drechsler, 2007). So verstehen Lezak, Howieson und Loring (2004) unter dem Konzept der Exekutivfunktionen Regulationsprozesse kognitiver, emotionaler und sozialer Funktionen. Als basale Komponenten nennen die Autoren *Volition* (Initiierung, Wahrnehmung, Motivation und Zielfindung), *Planung* (Objektivität und Voraussicht, Impulskontrolle, Realismus, Berücksichtigung von Alternativen), *zielorientiertes Handeln* (Organisation von Initiierung, Aufrechterhaltung, Switching und zeitlichen Aspekten) und *effektive Performanz* (Monitoring und Selbstkorrektur).

Auch Drechsler (2007) nimmt diesen Aspekt in ihre Taxonomie exekutiver Funktionen auf. Die Taxonomie ist insofern interessant, als sie unterschiedliche Beschreibungsebenen von Exekutivfunktionen voneinander abgrenzt, mit klinisch beobachtbaren Symptomen in Beziehung setzt und relevante Aufgabenparadigmen beschreibt. Drechsler (2007) liefert damit einen umfassenden Überblick zur Klassifikation von Störungen der Exekutivfunktionen. Die Autorin unterscheidet dabei drei Beschreibungsebenen, welche die Exekutivfunktionen hinsichtlich der zugrunde liegenden Prozesse (Initiierung/ Wechsel/ Inhibition), der Komplexität (basal/ komplex) sowie bezüglich verschiedener Regulationsebenen (Kognition/ Emotion/ Aktivität/ sozial) charakterisieren sollen (vgl. Abbildung 2.5)⁸.

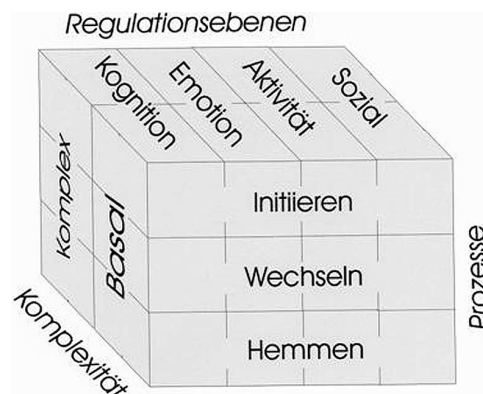


Abbildung 2.5: Beschreibungsebenen exekutiver Funktionen; aus Drechsler (2007, S. 236)

Insbesondere für die kognitive Regulationsebene spezifiziert Drechsler (2007) basale und komplexe Prozesse. Die basalen Prozesse (*Initiierung*, *Wechsel (shifting)*, *Inhibition* und das *Aufrechterhalten und Erneuern von Informationen aus dem Arbeitsspeicher*; vgl. Drechsler, 2007, S. 237) sowie

⁸Eine detaillierte Beschreibung aller Ebenen aus der Taxonomie einschließlich Informationen zu Störungsebenen und möglichen Aufgabenparadigmen findet sich in tabellarischer Form in Drechsler (2007, S. 237-238).

die komplexen Prozesse kognitiver Regulation (*Monitoring, Planen und Problemlösen, Organisation, Verteilung von Aufmerksamkeitsressourcen, Konfliktverarbeitung, Strategieranwendung*) decken sich dabei mit den beschriebenen Komponenten anderer Autoren (vgl. Repovš & Baddeley, 2006; Fuster, 2002; Miyake, 2000; Shallice et al., 1996; Smith & Jonides, 1999). Die Betrachtung der sozialen, emotionalen und Aktivitätsregulation weist jedoch auf darüber hinaus gehende Aspekte exekutiver Funktionen hin. So wird auf der Ebene der Aktivitätsregulation ein Fehlen von Aktivierung (Aktivitäts-hemmung) einem Fehlen von Hemmung (Aktivitätsüberschuss) gegenübergestellt. Die emotionale Regulation bezieht zum einen die Regulation von Affekten ein und zum anderen ein Lernen durch Feedback bzw. emotionale Bewertung. Bezüglich sozialer Regulationsprozesse wird zwischen einem Störungsbewusstsein die eigene Person betreffend (Introspektion) und dem Bewusstsein anderen Personen gegenüber (Aspekte der Empathie, der Perspektivübernahme, des Sozial- und Gesprächsverhaltens) unterschieden (Drechsler, 2007). Diese ganzheitliche Betrachtung ist insbesondere für Diagnostik und Therapie nach Hirnschädigung von Bedeutung, da nicht wenige neurologische Patienten auch Defizite in diesen nicht-kognitiven Komponenten exekutiver Funktionen aufweisen.

Den aufgeführten Theorien sind wesentliche Aspekte gemein. Als basale Komponenten können die Prozesse *Initiierung, Inhibition, Shifting* und *Monitoring* herausgestellt werden. Darüber hinaus betonen einige Autoren die zeitliche Integration als wichtigen kognitiv-exekutiven Prozess.

Alle genannten Komponenten sind auch bei der Lösung von Aufgaben zur Wortgenerierung von Bedeutung; Wortgenerierungsleistungen werden in der Literatur als Maß für Exekutivfunktionen beschrieben (Henry & Crawford, 2004; Kavé et al., 2011). Aufgrund ihrer zentralen Rolle im Rahmen dieser Arbeit folgt nun eine eingehende Betrachtung verschiedener theoretischer sowie neuroanatomischer Aspekte von Wortgenerierungsleistungen.

2.3 Wortgenerierungsfähigkeit

Zur Elizitierung der Wortgenerierungsfähigkeit (auch als *Wortflüssigkeit* oder *verbal fluency* bezeichnet) werden Versuchsteilnehmer aufgefordert, entsprechend einer vorgegebenen Regel (z.B. Wörter aus einer spezifischen lexikalischen oder semantischen Kategorie) in einer bestimmten Zeit so viele Nennungen wie möglich zu produzieren (Davis, 2007; Pekkala, 2012; Tröster, Warmflash, Osorio, Paolo, Alexander, & Barr, 1995; Unsworth, Spillers, & Brewer, 2011).

Die Evaluation der Wortgenerierungsfähigkeit wird häufig zur Einschätzung kognitiver und linguistischer Funktionen in verschiedenen Populationen eingesetzt (Davis, 2007; Pekkala, 2012; Ruff, Light, Parker, & Levin, 1997). So dient sie beispielsweise auch der Diagnose verschiedener neurologischer oder psychiatrischer Erkrankungen, wie Demenz (Araujo, Barca, Engedal, Coutinho, Deslandes, & Laks, 2011; Beatty, Testa, English, & Winn, 1997) und leichter kognitiver Beeinträchtigungen im Sinne eines *mild cognitive impairment* (Kleissendorf, Jaecks, & Stenneken, 2008; Nutter-Upham, Saykin, Rabin, Roth, Wishart, Pare, & Flashman, 2008; Östberg, Fernaeus, Hellström, Bogdanovic, & Wahlund, 2005), Schizophrenie (Gourovitch, Goldberg, & Weinberger, 1996; Paulsen, Romero, Chan, Davis, Heaton, & Jeste, 1996; Rossell, 2006), Depression (Araujo et al., 2011; Schmid, Strand, Ardal, Lund, & Hammar, 2011), Aphasie (Basso, Burgio, & Pradoni, 1997; Kim, Kim, Kim, & Heo, 2011), kommunikativ-kognitiver Beeinträchtigung (Bittner & Crowe, 2006; Gaspers, Thiele, Cimiano, Foltz, Stenneken, & Tscherepanow, 2012; Martin & McDonald, 2003), Morbus Parkinson (Piatt, Fields, Paolo, & Tröster, 1999b; Testa, Tröster, Fields, Gleason, Salmon, & Beatty, 1998), Chorea Huntington (Azambuja, Haddad, Radanovic, Barboso, & Mansur, 2007; Rosser & Hodges, 1994) sowie Multipler Sklerose (Beatty, 2002; Friend, Rabin, Groninger, Deluty, Bever, & Grattan, 1999; Henry & Beatty, 2006).

Die Wortgenerierungsfähigkeit ist insofern ein interessantes psychologisches Konstrukt, als es an der Schnittstelle sprachlicher und kognitiver Kompetenzen liegt. Neben der Integrität semantischer und lexikalischer Speichersysteme ist auch eine systematische Kontrolle des Wortabrufs, vor allem im Hinblick auf die vorgegebenen Regeln, wichtig. Die Bedeutung einzelner Prozesskomponenten, die für eine erfolgreiche Lösung einer Wortgenerierungsaufgabe erforderlich sind, werden zum Teil noch diskutiert und können sich auch in Abhängigkeit von der verwendeten Aufgabe unterscheiden (Pekkala, 2012; Ruff et al., 1997; Unsworth et al., 2011).

Im Folgenden werden zunächst die zur Elizitierung von Wortgenerierungsleistungen typischerweise verwendeten Aufgaben vorgestellt bevor anschließend zugrunde liegende Prozesskomponenten und neuroanatomische Zusammenhänge beschrieben werden.

2.3.1 Aufgabentypen

Es werden verschiedene Klassen von Wortgenerierungsaufgaben unterschieden (Aschenbrenner, Tucha, & Lange, 2000; Davis, 2007; Pekkala, 2012). Allen Aufgaben ist gemein, dass innerhalb eines begrenzten Zeitintervalls (meist 60, aber z.T. auch 30 oder 120 Sekunden) entsprechend einer vorgegebenen Regel so viele korrekte Nennungen wie möglich produziert werden

sollen (Aschenbrenner et al., 2000; Davis, 2007; Kim et al., 2011; Pekkala, 2012; Tröster et al., 1995). Die Wortgenerierungsaufgaben unterscheiden sich jedoch bezüglich der jeweiligen Regel, die bei der Wortproduktion beachtet werden soll. Im Wesentlichen werden zwei Klassen von Wortgenerierungsaufgaben voneinander abgegrenzt: lexikalische (auch: *lexical, letter, phonemic* oder *formal-based fluency*; vgl. Pekkala, 2012, Tröster et al., 1995) und semantische Aufgaben (auch: *semantic fluency, category fluency, category naming, category generation, controlled association task*; vgl. Davis, 2007, Grossman, 1981, Tröster et al. 1995). Weitere Differenzierungen betreffen die Komplexität der Aufgabenanforderung (Aschenbrenner et al., 2000; Henry & Phillips, 2006) und es werden darüber hinaus Aufgaben zur Verbgenerierung (auch: *action fluency*) beschrieben (Davis, Heidler-Gary, Gottesman, Crinion, Newhart, Moghekar, Soloman, Rigamonti, Cloutman, & Hillis, 2010; Delbeuck, Debachy, Pasquier, & Moroni, 2013; Pekkala, 2012; Piatt, Fields, Paolo, Koller, & Tröster, 1999a; Piatt, Fields, Paolo, & Tröster, 2004).

Bei lexikalischen Wortgenerierungsaufgaben sollen Wörter produziert werden, die mit demselben Anfangsbuchstaben beginnen, das heißt die Wortproduktion erfolgt innerhalb einer spezifischen lexikalischen Kategorie (Davis, 2007; Pekkala, 2012). Beispiele für simple lexikalische Wortgenerierungsaufgaben finden sich in Tabelle 2.1.

Tabelle 2.1: *Beispiele simpler lexikalischer Wortgenerierungsaufgaben*

Lexikalische Kategorie	Nennungen
Anfangsbuchstabe K	Kind, König, Krokodil, klatschen, krank, Kruste, ...
Anfangsbuchstabe B	Baum, Butter, baden, bunt, Blume, Bus, beten, ...
Anfangsbuchstabe M	Mich, munter, Mann, messen, Meter, Mantel, ...

Demgegenüber erfordern semantische Wortgenerierungsaufgaben das Produzieren von Wörtern innerhalb einer spezifischen semantischen Kategorie. Das heißt, es sollen Wörter produziert werden, die in einer bestimmten semantischen Relation zueinander stehen (Aschenbrenner et al., 2000; Pekkala, 2012). Beispiele für simple semantische Wortgenerierungsaufgaben finden sich in Tabelle 2.2.

Tabelle 2.2: *Beispiele simpler semantischer Wortgenerierungsaufgaben*

Semantische Kategorie	Nennungen
Tiere	Hund, Katze, Maus, Elefant, Papagei, Eule, ...
Gemüse	Tomaten, Paprika, Bohnen, Erbsen, Sellerie, ...
Kleidung	Hose, Rock, Hemd, Bluse, Kostüm, Pullover, ...

Zusätzlich zu den beschriebenen simplen lexikalischen und semantischen Wortgenerierungsaufgaben werden im Vergleich dazu Aufgaben mit komplexeren Anforderungen unterschieden (Aschenbrenner et al., 2000). Solche Aufgaben erfordern die alternierende Produktion von Wörtern aus zwei verschiedenen semantischen oder lexikalischen Kategorien und werden daher auch als Wechsel-Aufgaben (auch: *alternation fluency* oder *switching fluency*) bezeichnet (Baldo, Shimamura, Delis, Kramer, & Kaplan, 2001; Downes, Sharp, Costall, Sagar, & Howe, 1993; Henry & Phillips, 2006). Downes und Kollegen (1993) unterscheiden überdies zwischen intra- und extradimensionalen Wechsel-Aufgaben. Intradimensionale Wechsel-Aufgaben beinhalten den Wechsel zwischen Kategorien derselben Ebene (z.B. semantisch/semantisch) während extradimensionale Aufgaben den Wechsel zwischen Kategorien unterschiedlicher Ebenen (z.B. semantisch/lexikalisch) erfordern. Beispiele für die vergleichsweise häufig beschriebenen intradimensionalen lexikalischen und semantischen Wechsel-Aufgaben finden sich in der folgenden Tabelle⁹ (siehe Tabelle 2.3).

Tabelle 2.3: *Beispiele intradimensionaler lexikalischer und semantischer Wechsel-Aufgaben*

Wechsel-Kategorie	Nennungen
Lexikalisch F/A	Frost, Apfel, Fluss, Arbeit, fahren, ...
Semantisch Berufe/Obst	Bäcker, Apfel, Maurer, Birne, Metzger, ...

Eine weitere Klasse von Wortgenerierungsaufgaben, die in neuerer Zeit vermehrt bei der Evaluation von Wortgenerierungsleistungen Anwendung findet, ist die sogenannte *Verbgenerierung* (Davis et al., 2010; Delbeuck et al., 2013; Östberg et al., 2005; Pekkala, 2012; Piatt et al., 1999a,b, 2004). Bei der Verbgenerierung besteht die Aufgabe darin, in der vorgegebenen Zeit möglichst viele verschiedene Verben zu produzieren, also z.B. *schlafen, essen, riechen, schmecken, sehen, hören ...*. Eine typische Aufgabenstellung zur Elizitierung der Verbgenerierung ist beispielsweise (in Anlehnung an Piatt et al., 1999b): „Nennen sie Wörter die ausdrücken, was Lebewesen tun können.“

Die Bearbeitung der verschiedenen Aufgaben unterliegt spezifischen Regeln. Diese ergeben sich zum einen aus der verwendeten Aufgabe (lexikalische vs. semantisch, Verbgenerierung, simple vs. komplexe Anforderung), zum anderen werden in den Testinstruktionen weitere Regeln formuliert (Aschenbrenner et al., 2000; Beatty, 2002; Ruff et al., 1997). So wird bei allen Aufgaben explizit darauf hingewiesen, dass keine Wörter mehrfach ge-

⁹Der Schrägstrich (/) trennt bei den beiden Wechsel-Aufgaben jeweils die lexikalischen bzw. semantischen Kategorien, zwischen denen bei der Wortgenerierung alterniert werden soll.

nannt werden dürfen. Bei den lexikalischen Aufgaben sowie bei der Aufgabe zur Verbgenerierung schließt das auch Wörter mit bedeutungsgleichem Wortstamm (z.B. *Mann, Männer ...* oder *essen, aß, ...*) ein. Darüber hinaus dürfen keine Eigennamen, wie z.B. *Martina, München, Melitta ...*, genannt werden (Aschenbrenner et al., 2000).

2.3.2 Funktionen & neuroanatomische Korrelate

Die Fähigkeit zur Wortgenerierung unterliegt sowohl lexikalischen und semantischen Verarbeitungsprozessen als auch einer kognitiv-exekutiven Überwachung dieser Prozesse (Unsworth et al., 2011). Dabei scheinen die Leistungen in den verschiedenen Wortgenerierungsaufgaben auf sich überlappenden aber auch auf unterschiedlichen Prozessen zu beruhen (Tröster et al., 1995).

Für alle Wortgenerierungsaufgaben wird generell eine Beteiligung von Exekutivfunktionen angenommen. Dabei werden spezifische Komponenten herausgestellt. Eine besondere Rolle spielen nach Meinung verschiedener Autoren (i) die *Initiierung* des Wortabrufprozesses inklusive des Bereitstellens von Aufmerksamkeitsmechanismen, (ii) ein kontinuierliches *Monitoring* der Aufgabe, (iii) die *Inhibition* bereits genannter Wörter, sowie (iv) die *Anwendung von Strategien* welche auch die Fähigkeit zum *set-shifting*, das heißt dem Wechsel zwischen Strategien oder Konzepten, beinhaltet (Davis, 2007; Drechsler, 2007; Henry & Beatty, 2006; Henry & Phillips, 2006; Pekkala, 2012; Rosen & Engle, 1997; Rosser & Hodges, 1994; Ruff et al., 1997; Tröster et al., 1995; Unsworth et al., 2011).

Diese kognitiv-exekutiven Prozesse operieren auf lexikalischem und semantischem Wissen. Die Lösung semantischer Wortgenerierungsaufgaben beruht auf der Integrität des semantischen Systems bzw. semantischer Netzwerke (Pekkala, 2012; Tröster et al., 1995). Eine erfolgreiche semantische Wortgenerierung operiert auf semantischen Verarbeitungsmechanismen und bedient sich semantischer Suchstrategien. Neben intakten semantischen Repräsentationen ist hier die Fähigkeit zur Kategorisierung auf Basis semantischer Merkmale erforderlich (Pekkala, 2012). Demgegenüber erfordert die lexikalische Wortgenerierung den Abruf einer korrekten Wortform und bedient sich einer lexikalischen Suchstrategie (Davis, 2007). Es wird beschrieben, dass der semantischen Wortgenerierung ein eher assoziativ automatischer und der lexikalischen Wortgenerierung ein eher strategischer Prozess zugrunde liegt (vgl. Unsworth et al., 2011). Die Lösung einer lexikalischen Wortgenerierungsaufgabe erfordert demnach, alltagserprobte semantische zugunsten lexikalischer Suchstrategien aufzugeben (Aschenbrenner et al., 2000; Crowe, 1992; Henry & Crawford, 2004). Das betonen auch Basso und Kollegen (1997). Die Autoren argumentieren, dass eine semantische Suchstrategie entsprechend der Organisation des semantischen Systems der Stra-

ategie entspricht, die auch in der normalen Kommunikation beim Wortabruf verwendet wird, während eine lexikalische oder phonologische Strategie eher unüblich ist (Basso et al., 1997). Lexikalische Wortgenerierungsaufgaben erfordern daher mehr Kapazität für die Initiierung und Aufrechterhaltung systematischer Abrufstrategien und werden somit im Vergleich zu semantischen Aufgaben als schwieriger und exekutiver beschrieben (Basso et al., 1997; Piatt et al., 1999a; Rosser & Hodges, 1994). Was jedoch nicht bedeutet, dass der semantischen Wortgenerierung keine strategische Komponente zugrunde liegt (Unsworth et al., 2011).

Lexikalische und semantische Wortgenerierungsfähigkeit werden in der Literatur als dissoziierbare Leistungen beschrieben (Baldo, Schwartz, Wilkins, & Dronkers, 2010). Während lexikalische Wortgenerierungsleistungen stärker in Verbindung mit exekutiven Funktionen des frontalen Kortex beschrieben werden, werden semantische Wortgenerierungsleistungen mit dem temporalen Kortex und dem semantischen System in Verbindung gebracht (Baldo et al., 2001, 2010; Henry & Crawford, 2004; Pekkala, 2012). Bildgebungsstudien konnten jedoch auch zeigen, dass sowohl bei lexikalischen als auch bei semantischen Wortgenerierungsleistungen Teile des Frontalhirns aktiviert werden (Pekkala, 2012). Unsworth und Kollegen (2011) führen das auf die Annahme zurück, dass jedem Abrufprozess generell eine strategische Komponente zugrunde liegt. Darüber hinaus werden für die beiden Klassen von Wortgenerierungsaufgaben distinkte Aktivierungsmuster beschrieben, was auf eine Beteiligung verschiedener Wortabruf- bzw. -speichersysteme (semantisch/konzeptuell vs. phonologisch/lexikalisch) zurückgeführt wird (Pekkala, 2012). Beeinträchtigungen der lexikalischen Wortgenerierungsleistungen werden vor allem im Zusammenhang mit dem rechten und linken präfrontalen Kortex, sowie mit dem Gyrus Cinguli, und dem Cerebellum (rechts) beschrieben (Pekkala, 2012). Semantische Wortgenerierungsleistungen sind demgegenüber zwar auch mit Aktivierung im linken und rechten frontalen Kortex assoziiert, werden darüber hinaus aber insbesondere mit Aktivierung in weiten Teilen des linken, medialen Temporallappens in Verbindung gebracht (Pekkala, 2012). Evidenz hierfür ergibt sich vor allem auch daraus, dass Patienten mit Läsionen des Frontallappens schlechtere Leistungen in lexikalischen Wortgenerierungsleistungen zeigen, während bei Patienten mit Läsionen im Temporallappen Defizite bei der semantischen Wortgenerierung überwiegen (Troyer, Moscovitch, Winocur, Alexander, & Stuss, 1998; Henry & Crawford, 2004; Kavé et al., 2011; Rosser & Hodges, 1994). Generell sollen links frontale Läsionen im Vergleich zu rechts frontalen Läsionen mit schlechteren Wortgenerierungsleistungen assoziiert sein, was als Hinweis darauf gewertet wird, dass die Wortgenerierungsfähigkeit vor allem links frontal repräsentiert zu sein scheint (Baldo et al., 2001; Lezak et al., 2004).

Eine relative Prominenz von Leistungsdefiziten in lexikalischen gegenüber semantischen Wortgenerierungsaufgaben oder vice versa kann Aussagen darüber ermöglichen, ob bei einem spezifischen Störungsbild primär ein Defizit in den Exekutivfunktionen oder der Semantik vorliegt (Henry & Beatty, 2006; Henry & Phillips, 2006). Dabei weisen schlechtere Leistungen in semantischen Wortgenerierungsaufgaben auf ein Defizit in semantischen Verarbeitungsprozessen hin, während ähnliche Leistungen in semantischen und lexikalischen Wortgenerierungsaufgaben sowie schlechtere Leistungen in lexikalischen im Vergleich zu semantischen Wortgenerierungsaufgaben ein Defizit in den Exekutivfunktionen bzw. bei der lexikalischen Verarbeitung nahelegen (Henry & Beatty, 2006). Dies lässt sich mit der Annahme zwei verschiedener Wortspeichersysteme begründen, das heißt mit der Annahme, dass Wortbedeutung und Wortform voneinander unabhängig repräsentiert sind (Basso et al., 1997). Sehr vereinfacht gesprochen ist die Bedeutung eines Wortes im semantischen System repräsentiert. Ausgehend von der Wortbedeutung erfolgt eine Aktivierung der zugehörigen Wortform im Lexikon, bevor das Wort produziert wird. Nach dieser modelltheoretischen Vorstellung führen Störungen im semantischen System zu beeinträchtigten semantischen Wortgenerierungsleistungen, während Störungen, die das Lexikon betreffen sich in Form verminderter lexikalischer und semantischer Wortgenerierungsleistungen zeigen können (Basso et al., 1997).

Die Wechsel-Aufgaben gelten im Vergleich zu den simplen Wortgenerierungsaufgaben als kognitiv-exekutiv anspruchsvoller, da sie in besonderem Maße *set-shifting* Funktionen erfordern (Henry & Phillips, 2006; Keil & Kaszniak, 2002; Nutter-Upham et al., 2008). Die alternierende Produktion von Wörtern aus zwei verschiedenen Kategorien erfordert ein hohes Maß an kognitiver Flexibilität, Monitoring sowie eine zeitliche Integration von Initiierung und Inhibition (Henry & Phillips, 2006). Obwohl es Befunde gibt, die zeigen, dass sich Patienten verschiedener Populationen durch eine Evaluation von Leistungen in Wechsel-Aufgaben von Sprachgesunden dissoziieren lassen (siehe z.B. Gourovitch et al. 1996 für Schizophrenie oder Nutter-Upham et al., 2008 für *mild cognitive impairment*), gibt es derzeit nur wenige Untersuchungen von Wortgenerierungsleistungen, die eine Evaluation von Wechsel-Aufgaben berücksichtigen.

Leistungen bei der Verbgenerierung werden im Zusammenhang mit fronto-subkortikalen Schaltkreisen diskutiert. Defizite beim Produzieren von Verben werden vor allem mit Läsionen im frontalen Kortex sowie in subkortikalen Arealen in Verbindung gebracht, während die Produktion von Nomen eher im temporalen Kortex repräsentiert zu sein scheint (Davis et al., 2010). Piatt und Kollegen (1999a; 1999b) fanden einen Zusammenhang zwischen fronto-striataler Pathologie und Verbgenerierungsleistungen bei Patienten mit Morbus Parkinson und konnten anhand der Leistung bei der Verbgene-

rierung Patienten mit Morbus Parkinson mit und ohne Demenz voneinander dissoziieren. Die Autoren interpretierten die Verbgenerierungsperformanz als Maß für Leistungen in den Exekutivfunktionen.

In Bezug auf die Verbgenerierung bestehen noch Kontroversen bezüglich der Relevanz syntaktischer Verarbeitung für die Performanz in Verbgenerierungsaufgaben. Vor allem im Zusammenhang mit der Verbvalenz wird diskutiert, welchen Anteil frontal repräsentierte syntaktische Kompetenzen an der Verbgenerierungsleistung haben (Östberg et al., 2007; Pekkala, 2012).

Zusammengenommen lässt sich festhalten, dass die Wortgenerierung neben intakten lexikalischen und semantischen Prozessen der Speicherung und Enkodierung sowie des Abrufs auch kognitiv-exekutive Funktionen erfordert, welche für die zeitliche Integration und einen kontrollierten Ablauf des Wortgenerierungsprozesses zuständig sind. Als beteiligte Komponenten exekutiver Funktionen gelten *Initiierung*, *Monitoring*, *Inhibition* sowie die Anwendung von Strategien inklusive (*set*)-*shifting* (Henry & Beatty, 2006; Henry & Phillips, 2006; Pekkala, 2012; Rosen & Engle, 1997; Rosser & Hodges, 1994; Ruff et al., 1997; Tröster et al., 1995; Unsworth et al., 2011). Die Evaluation von Wortgenerierungsleistungen gilt generell als sensitives Maß zur Erfassung kognitiver Leistungen und linguistischer Funktionen bei Patienten mit Hirnschädigung und ermöglicht Aussagen über Wortabruf/-produktion und Leistungen in den Exekutivfunktionen (Henry & Crawford, 2004). Weitere Informationen zur Evaluation von Wortgenerierungsleistungen und Verfahren, die hierzu eingesetzt werden, folgen bei der Besprechung diagnostischer Aspekte in Abschnitt 2.5.1.

2.4 Sprach- und Kommunikationsstörungen

Im Rahmen dieser Dissertation steht die Evaluation von Wortgenerierungsleistungen bei verschiedenen klinischen Populationen im Vordergrund. Untersucht werden Patienten mit sprachsystematischen Beeinträchtigungen im Sinne einer Aphasie, Patienten mit kommunikativ-kognitiven Defiziten sowie Patienten mit Multipler Sklerose unterschiedlicher Verlaufsformen. An dieser Stelle sollen nun einige theoretische Grundlagen zu diesen Störungsbilder geliefert werden. Insbesondere eine terminologische Abgrenzung ist an dieser Stelle wichtig, da die Grenzen zwischen den verschiedenen Formen beobachtbarer Beeinträchtigungen sowie ihre Zuordnung zum Teil nicht einheitlich sind. So werden beispielsweise abhängig von der Ätiologie¹⁰ oder der Lokalisation einer Hirnschädigung Sprach- bzw. Kommunikationsstörungen unterschieden bzw. spezifische Symptome oder auch Symptomkombinationen mit einer spezifischen Ätiologie oder Lokalisation in Verbindung

¹⁰Der Begriff Ätiologie stammt aus dem Griechischen (*aitía*) und wird zur Bezeichnung der Ursache einer Krankheit verwendet (Psyrembel, 2002).

gebracht (Huber, Poeck, & Weniger, 2006; Wehmeyer & Grötzbach, 2006; Weniger, 2006).

2.4.1 Sprachsystematische Defizite: Aphasie

Aphasien¹¹ werden als erworbene, zentrale Sprachstörung beschrieben, die nach abgeschlossenem Spracherwerb auftreten (Huber et al., 2006). Dabei können sowohl rezeptive als auch produktive Fähigkeiten der Laut- und Schriftsprache betroffen sein. Aphasien manifestieren sich zumeist auf allen Ebenen des Sprachsystems (Phonologie, Lexikon, Semantik, Syntax, Pragmatik) und werden daher auch als multi- bzw. supramodale Störungen bezeichnet (Huber et al., 2006). Die jeweilige Ausprägung sowie der Schweregrad können in Abhängigkeit von Lokalisation, Größe und Art der Hirnschädigung interindividuell variieren (Huber et al., 2006; Weniger, 2006; Wehmeyer & Grötzbach, 2006). Typische Symptome sind beispielsweise Wortfindungs- bzw. Wortabrufstörungen, phonematische oder semantische Paraphasien, ein Fehlen von Funktionswörtern und/oder eine falsche Verwendung von Flexionsformen, Satzabbrüche aber auch Satzteilverdopplungen oder Satzverschränkungen, das Auftreten von sprachlichen Stereotypen oder Sprachautomatismen sowie Sprachverständnisstörungen, welche sich mehr oder weniger stark auf die kommunikative Interaktion des Betroffenen auswirken können. Die beobachtbaren Defizite sind insgesamt sprachstruktureller bzw. sprachsystematischer Natur (Huber et al., 2006).

Diese für eine Aphasie typischen Symptome treten jedoch nicht in willkürlicher Kombination auf. Es lassen sich vielmehr, insbesondere bei vaskulärer Ätiologie und nach Ende der Akutphase, spezifische Muster an Symptomkombinationen erkennen, aus welchen sich die vier Standardsyndrome der Aphasien ableiten lassen (Huber et al., 2006; Weniger, 2006). Unterschieden werden die Amnestische Aphasie (vgl. Poeck, Kerschensteiner, Stachowiak, & Huber, 1974), die Broca-Aphasie (vgl. Kerschensteiner, Poeck, Huber, Stachowiak, & Weniger, 1978), die Wernicke-Aphasie (vgl. Huber, Stachowiak, Poeck, & Kerschensteiner, 1975) sowie die Globale Aphasie (vgl. Stachowiak, Huber, Kerschensteiner, Poeck, & Weniger, 1977), welche jeweils durch spezifische Kardinalsymptome charakterisiert sind. Darüber hinaus werden als sogenannte Nicht-Standardsyndrome die Leitungsaphasie sowie die Transkortikalen Aphasien beschrieben, bei denen modalitätsspezifische Charakteristika im Vordergrund stehen (Huber et al., 2006). Für einen detaillierten Überblick zu den Standard- und Nicht-Standardsyndromen siehe Huber et al. (2006) oder Weniger (2006). Von besonderer Relevanz ist auch die Restaphasie, bei der aphasische Symptome in so geringer Ausprägung auftreten, dass sie mit Standardtestverfahren nicht verlässlich identifiziert

¹¹Der Begriff Aphasie stammt aus dem Griechischen (*aphasia*) und bedeutet übersetzt soviel wie *ohne Sprache* (Wehmeyer & Grötzbach, 2006)

werden können. Betroffene Patienten berichten aber dennoch von sprachlichen Einschränkungen, berufliche Wiedereingliederung kann erschwert und Partizipation beeinträchtigt sein. Eine detaillierte Übersicht zur Restaphasie liefert Jaecks (2006). Weitere Einteilungen der Aphasie orientieren sich darüber hinaus an der Flüssigkeit der Sprachproduktion (flüssig vs. nicht flüssig), der Lokalisation (kortikal vs. subkortikal), der Krankheitsdauer (akut, postakut, chronisch) oder der Ätiologie (Huber et al., 2006; Jaecks, 2006; Wehmeyer & Grötzbach, 2006; Weniger, 2006).

Sprachsystematische Beeinträchtigungen im Sinne einer Aphasie sind nach kortikaler Schädigung der sprachdominanten, meist linken, Hemisphäre zu beobachten. Innerhalb der sprachdominanten Hemisphäre lassen sich sprachrelevante Regionen identifizieren. Diese sind in perisylvischen Arealen lokalisiert, erstrecken sich somit über die an die Sylvische Furche angrenzenden Gyri des Frontal, Temporal- und Parietallappens (Huber et al., 2006; Poeck & Hacke, 2001). Als häufigste Ätiologie einer Aphasie sind zerebrale Gefäßinsulte, das heißt zerebro-vaskuläre Durchblutungsstörungen (Schlaganfälle) im Versorgungsgebiet der Arteria cerebri media zu nennen. Als weitere Ursachen werden Schädelhirntraumata, Hirntumoren, Enzephalitis sowie Hirnatrophie genannt (Huber et al., 2006; Wehmeyer & Grötzbach, 2006). In Abbildung 2.6 findet sich eine Darstellung relevanter Sprachregionen des menschlichen Gehirns sowie des Versorgungsgebiets der Arteria cerebri media.

Davon abzugrenzen sind Sprachstörungen subkortikaler Genese (Crosson, 1999, 2012; Heidler, 2009; Hillis, Wityk, Barker, Beauchamp, Gailloud, Murphy, Cooper, & Metter, 2002; Kuljic-Obradovic, 2003; Nadeau & Crosson, 1997). Als relevant werden insbesondere Läsionen des linksseitigen Thalamus sowie der linksseitigen Basalganglien hervorgehoben, aber auch Schädigungen der Capsula Interna sowie der weißen Substanz und des subkortikalen Marklager beschrieben (Crosson, 1985, 1999, 2012; Kuljic-Obradovic, 2003; Wallesch, 2006). Dem Thalamus kommt dabei eine besondere Rolle bei der lexikalisch-semantischen Verarbeitung zu (Crosson, 1999; Kuljic-Obradovic, 2003) und wird darüber hinaus zusammen mit den Basalganglien als wichtige Struktur bei der Integration und Regulation sprachlicher Prozesse genannt (Nadeau & Crosson, 1997). Darüber hinaus wird die Bedeutung kortiko-subkortikaler Konnektionen für die Sprachverarbeitung betont. Zu Sprachstörungen subkortikaler Genese kann es unter anderem durch Schädigung der Strukturen selbst oder aber durch eine funktionale Diskonnektion von Verbindungen zwischen subkortikalen Strukturen und kortikalen Spracharealen kommen (Heidler, 2009; Hillis et al., 2002; Nadeau & Crosson, 1997). Im Gegensatz zu kortikalen Aphasien wird Sprachstörungen subkortikaler Genese oftmals ein distinkter Pathomechanismus zugeschrieben. So sollen Sprachstörungen nach Schädigung subkortikaler Strukturen sekun-

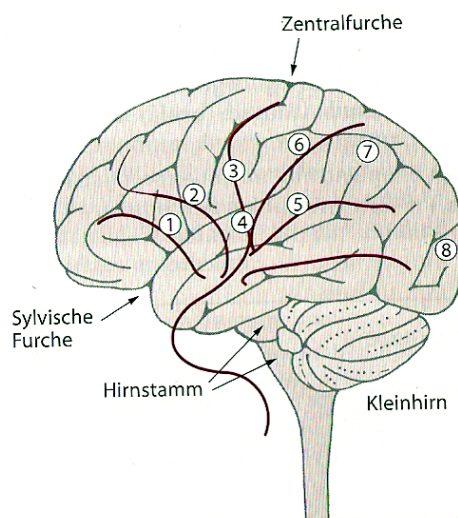


Abbildung 2.6: Sprachregionen im menschlichen Gehirn. 1 Broca-Area; 2 motorische Gesichtsregion; 3 somatosensorische Gesichtsregion; 4 Hörfelder, 5 Wernicke-Area; 6 Gyrus supramarginalis; 7 Gyrus angularis; 8 visuelle Assoziationsregion; rot Äste der A. cerebri media (W. Huber, Aachen); aus Poeck & Hacke (2001, S. 167)

där als Folge beeinträchtigter kognitiver Funktionen entstehen, insbesondere bedingt durch Beeinträchtigungen in den Exekutivfunktionen (Heidler, 2009). Es werden jedoch auch Hinweise darauf beschrieben, dass subkortikale Schädigungen von kortikalen Läsionen begleitet sein können und auftretende Sprachstörungen durch eine Minderperfusion kortikaler Areale entstehen (vgl. hierzu Hillis et al., 2002 oder Wallesch, 2006). Die Symptome subkortikaler Sprachstörungen weisen eine große Heterogenität auf, was unter anderem auf die komplexe Verschaltung kortikaler und subkortikaler Strukturen zurückgeführt wird (Heidler, 2009). Dabei unterscheiden sich die auftretenden Symptomkombinationen von denen kortikaler Aphasien. Während bei kortikalen Aphasien von einer multimodalen Störung ausgegangen werden kann, lassen sich bei subkortikalen Sprachstörungen modalitätsspezifisch intakte Leistungen erkennen. Hervorzuheben sind hier die Leistungen beim Nachsprechen, aber auch Leistungen in der Schriftsprache und auf Ebene der Syntax gelten als gut erhalten (Heidler, 2009; Kuljic-Obradovic, 2003).

Begriffe wie *subkortikale Aphasie* (Kuljic-Obradovic, 2003; Nadeau & Crosson, 1997) oder *thalamische Aphasie* (Crosson, 2012; Heidler, 2009) werden häufig zur Bezeichnung von Sprachstörungen subkortikaler Genese verwendet. Die Verwendung des Begriffs *Aphasie* ist jedoch nicht immer wirklich gerechtfertigt. Bis heute wird der Pathomechanismus hinter diesen Störungen kontrovers diskutiert. Es ist nicht abschließend geklärt, welchen

Anteil primär kognitive Defizite bzw. Defizite bei der Integration kognitiver Funktionen bei der Entstehung von Sprachstörungen nach subkortikalen Läsionen haben und in welchem Ausmaß beispielsweise eine Minderperfusion kortikaler Areale nach subkortikaler Schädigung zu primären Sprachstörungen führen kann (Wallesch, 2006). Auf die Verwendung des Begriffs *Aphasie* sollte daher verzichtet werden, wenn nicht primär sprachsystematische Leistungen betroffen sind, sondern es wahrscheinlich scheint, dass sprachliche Symptome sekundär durch Beeinträchtigung kognitiver Funktionen hervorgerufen wurden. Die alternative Verwendung eines eher beschreibenden Terminus wie etwa *kognitiv-linguistisches* oder *kommunikativ-kognitives Defizit* scheint hier angemessener (vgl. hierzu auch 2.4.2). Wird im Folgenden von *Aphasien* gesprochen sind sprachsystematische Beeinträchtigungen kortikaler Genese gemeint.

Wortgenerierungsleistungen bei Aphasie

Wie in Abschnitt 2.3 beschrieben, erfordert die Wortgenerierung eine intakte lexikalische und semantische, sprachsystematische Verarbeitung. Einschränkungen der Wortgenerierungsleistungen sind daher im Rahmen aphasischer Sprachstörungen zu erwarten (Adams, Reich & Flowers, 1989; Basso et al., 1997; Davis, 2007; Grossman, 1981). Die Überprüfung der Wortgenerierungsfähigkeit ist daher häufig auch Bestandteil verschiedener Testverfahren zur Aphasiediagnostik, z.B. in der *Boston Diagnostic Aphasia Examination* (BDAE; Goodglass & Kaplan, 1972) oder dem *Bielefelder Aphasie Screening* (BIAS; Richter, Wittler & Hielscher-Fastabend, 2006), insbesondere um auch leichte bis subtile Defizite identifizieren zu können (Richter et al., 2006).

Einschränkungen der Wortgenerierungsleistungen bei Aphasie werden in Form einer geringeren *Anzahl korrekter Nennungen* berichtet (Basso et al., 1997; Davis, 2007). Nach Basso und Kollegen (1997) zeichnet sich die Wortgenerierungsleistung bei Aphasie insbesondere dadurch aus, dass aphasische Patienten insgesamt weniger Nennungen produzieren; das heißt, die geringere *Anzahl korrekter Nennungen* ergibt sich nicht dadurch, dass Aphasiker mehr falsche Nennungen produzieren (Basso et al., 1997). Befunde zu Leistungsunterschieden bei der lexikalischen und semantischen Wortgenerierung weisen darauf hin, dass lexikalische Wortgenerierungsleistungen bei aphasischen im Vergleich zu nicht-aphasischen Patienten mit Hirnschädigung stärker beeinträchtigt sind (Basso et al., 1997). So konnten Basso und Kollegen (1997) zeigen, dass viele Aphasiker adäquate Leistungen in der semantischen und beeinträchtigte Leistungen bei der lexikalischen Wortgenerierung zeigen. Semantische, nicht aber lexikalische Wortgenerierungskompetenz korrelierte dabei mit Leistungen im konfrontativen Benennen. Ergebnisse aus einer Untersuchung von Baldo und Kollegen (2010) weisen daraufhin, dass sich

auch bei Aphasien insbesondere die Läsionslokalisation auf die Manifestation der Leistungen in semantischen und lexikalischen Wortgenerierungsaufgaben auswirkt. Bei zwei Patienten mit chronischer Aphasie konnten die Autoren Defizite in lexikalischen Wortgenerierungsleistungen nach links frontaler Läsion und Defizite bei der semantischen Wortgenerierung nach links temporalen Läsion dissoziieren.

Viele Studien zur Wortgenerierung bei Aphasie konzentrieren sich ausschließlich auf die Evaluation semantischer Wortgenerierungsleistungen. So konnten Kim und Kollegen (2011) bei einer Analyse semantischer Wortgenerierungsperformanz zeigen, dass Patienten mit Aphasie häufig nur in den ersten 30 Sekunden einer Aufgabe überhaupt Nennungen produzieren. Grossman (1981) fand, dass Patienten mit flüssiger und nicht-flüssiger Aphasie im Vergleich zu Patienten ohne Aphasie nach rechthemisphärischen Läsionen sowie gegenüber Sprachgesunden signifikant weniger korrekte Nennungen generierten. Unterschiede in den Wortgenerierungsleistungen zwischen Patienten mit flüssiger und nicht-flüssiger Aphasie zeigten sich nicht. Eine qualitative Analyse der Wortgenerierungsleistungen konnte zudem zeigen, dass Patienten mit leichter Aphasie nicht nur eine geringere Anzahl weniger gebräuchlicher Vertreter einer Kategorie produzierten, sondern auch Unterschiede bei der Generierung typischer Vertreter zu finden waren¹² (Adams et al., 1989).

Die Ergebnisse aus den Studien weisen darauf hin, dass die *Anzahl korrekter Nennungen* ein gutes Leistungsmaß zu sein scheint, um zwischen Patienten mit Aphasien und anderen Störungsbildern bzw. Sprachgesunden zu dissoziieren, nicht aber für eine Abgrenzung verschiedener Subtypen einer Aphasie geeignet ist (Grossman, 1981).

2.4.2 Kommunikativ-kognitive Defizite

Der Terminus *kommunikativ-kognitives Defizit* wird zur Bezeichnung von Störungen in der Kommunikation auf Ebene der Pragmatik verwendet, welche durch eine Beeinträchtigung des komplexen Zusammenspiels kognitiver und sprachlicher Funktionen entstehen (Coelho & DeRuyter, 1996). Ihre primäre Ursache wird in einer Störung kognitiver Funktionen gesehen, welche bei der Integration von sprachlichem und nicht-sprachlichem Wissen für eine erfolgreiche Kommunikation erforderlich sind (Coelho & DeRuyter, 1996; Cummings, 2007; Lê et al., 2011). Die beeinträchtigten kognitiven Funktionen reichen dabei von relativ basalen bis hin zu komplexeren Leis-

¹²Ein Beispiel für einen typischen Vertreter aus der Kategorie *Tiere* wäre z.B. *Hund*, während weniger gebräuchliche Vertreter beispielsweise *Gürteltier* oder *Würfelfalle* wären.

tungen und umfassen neben Exekutivfunktionen auch Gedächtnis- und Aufmerksamkeitsfunktionen. Während sprachstrukturelle Leistungen bei Patienten mit kommunikativ-kognitiven Defiziten als unbeeinträchtigt gelten, ist die sozial-kommunikative Verwendung von Sprache beeinträchtigt. Lê und Kollegen (2011) bezeichnen Patienten mit kommunikativ-kognitiven Defiziten als "good talker but bad communicator" (Lê et al., 2011, S. 4).

Die klinischen Symptome kommunikativ-kognitiver Beeinträchtigungen sind sehr heterogen. Es werden vor allem Defizite in Diskurs und Konversation beschrieben, das heißt bei der Verwendung von Sprache in der sozialen Interaktion (Cummings, 2007; Lehman Blake & Tompkins, 2009). Dazu gehören beispielsweise Schwierigkeiten bei der Verarbeitung non-literaler Sprache (z.B. Sarkasmus, Humor, Ironie), welche sich in der literalen Interpretation entsprechender Äußerungen zeigen, oder auch Defizite bei der Disambiguierung mehrdeutiger Ausdrücke und Äußerungen. Die Konversation beinhaltet häufig nebensächliche Details und wird darüber hinaus als egozentrisch und wenig fokussiert mit schnellen Themenwechseln beschrieben. Der Informationsgehalt der Äußerungen gilt als reduziert und es lässt sich ein Mangel an Kohärenz und Kohäsion im Diskurs feststellen. Patienten mit kommunikativ-kognitiven Defiziten haben Schwierigkeiten beim Ziehen von Inferenzen; Intentionen des Gesprächspartners werden nicht beachtet oder fehlinterpretiert. Darüber hinaus lässt sich ein latenter Mangel an Inhibition sozial nicht angemessener Äußerungen feststellen, der sich im Auftreten ebendieser Äußerungen manifestiert. Hinsichtlich quantitativer Aspekte werden sowohl erheblich reduzierte bis hin zu extrem ausschweifende Äußerungen berichtet (vgl. Lehman Blake, 2005; Lehman Blake & Tompkins, 2009; Cummings, 2007; Martin & McDonald, 2003).

Häufig wird in diesem Zusammenhang auch von Defiziten der *high-level language* gesprochen (Laakso, Brunnegård, Hartelius, & Ahlsén, 2000). Wong und Kollegen (2010) haben analog zur Definition von kognitiven "low vs. high-level" Kompetenzen Definitionen in Bezug auf sprachliche Kompetenzen formuliert (in Lindfors, 1987). Demnach erfordern "low-level"-Funktionen das Wiedergeben zuvor erhaltener Informationen, wie z.B. beim Nachsprechen oder beim Benennen, während "high-level" Sprachfunktionen darüber hinaus gehen und beispielsweise eine Re-Organisation, ein Ziehen von Inferenzen, ein in Beziehung setzen oder Problemlösen erfordern. *High-level language* Aufgaben zeichnen sich demnach durch eine ausgeprägte kognitiv-exekutive Komponente aus. Es handelt sich per Definition um Aufgaben, deren Lösung die Unterstützung frontaler Hirnareale, das heißt auch Exekutivfunktionen, erfordert.

Zur Bezeichnung ebendieser kommunikativ-kognitiven Defizite existieren darüber hinaus weitere Termini, die meist synonym verwendet werden. Dazu zählen *nonaphasic communication deficits* oder *non-aphasic language*

disorder (McDonald, 1993; McDonald & Pearce, 1998), *Kognitive Dysphasie* (Heidler, 2010), *pragmatic language disorder* (McDonald, 2000; Martin & McDonald, 2003) oder *pragmatic aphasia* (Myers, 2001), *right hemisphere damage-syndrom* (auch *RHD-syndrom*, Myers, 2001) oder *high-level cognitive-linguistic deficit* (Wong et al., 2010).

Kommunikativ-kognitive Störungen sind mit Standardverfahren und -prozeduren, z.B. aus der Aphasiediagnostik, nicht reliabel zu erfassen (Barrow, Hough, Rastatter, Walker, Holbert & Rotondo, 2003; Wong et al., 2010). Das Störungsbild ist insbesondere dadurch charakterisiert, dass Leistungen in strukturierten Settings häufig unauffällig sind und erst im Alltag bzw. in unstrukturierten Situationen mit komplexen Anforderungen Beeinträchtigungen erkennbar werden (Coelho et al., 2005; Turkstra et al., 2005a). Doch sogar subtile Defizite können sich auf die soziale Teilhabe auswirken und eine erfolgreiche Rückkehr in den Beruf behindern (Duff & Proctor, 2002; Wong et al., 2010). Etwa ein Drittel der Patienten mit kommunikativ-kognitiven Defiziten leiden unter anhaltenden Beeinträchtigungen, die sie in ihrem täglichen Leben einschränken (Coelho & DeRuyter, 1996).

Kommunikativ-kognitive Defizite werden nach Hirnschädigungen verschiedener Ätiologien und Lokalisationen beschrieben. Insbesondere nach zerebraler Schädigung der rechten Hemisphäre oder im Bereich des (prä)frontalen Kortex wird von kommunikativ-kognitiven Beeinträchtigungen berichtet (Blake, Duffy, Myers, & Tompkins, 2002; Cummings, 2007; Martin & McDonald, 2003). Aber auch subkortikale Läsionen, wie in Abschnitt 2.4.1 beschrieben, können kommunikativ-kognitive Defizite hervorrufen. Zu den häufigsten Ätiologien werden neben Schädelhirntraumata¹³ auch zerebrale Gefäßinsulte, Schizophrenien, verschiedene neurodegenerative Erkrankungen oder Autismus gezählt (Cummings, 2007; Martin & McDonald, 2003). Kommunikativ-kognitive Defizite können sich, wie etwa bei diffuser axonaler Schädigung nach einem minimalen Schädelhirntrauma, auch sehr subtil äußern, sind dann schwer greif- und messbar und werden häufig nur dadurch erkannt, dass sie zu Problemen im Alltag führen (Russell, Arenth, Scanlon, Kessler, & Ricker, 2012).

Es existieren verschiedene Theorien zur Erklärung kommunikativ-kognitiver Defizite, die vor allem auf Annahmen zu spezifischen Ätiologien zurückzuführen sind (Martin & McDonald, 2003). So werden Defizite nach rechtshemisphärischen Läsionen häufig auf eine Störung in der *Theory of Mind* (*Social inference theory*) zurückgeführt. Demgegenüber besteht bei Autismus unter anderem die Annahme einer generell reduzierten zentralen

¹³Schädelhirntraumata weisen meist eine Schädigung des Frontalhirns und damit verbundener Strukturen auf (Martin & McDonald, 2003)

Kohärenz, das heißt der situative Kontext kann nicht im erforderlichen Maß integriert werden, was zu den beobachtbaren Defiziten führt (*Weak central coherence hypothesis*). Eine weitere Theorie, die sich insbesondere auf Beobachtungen bei Patienten mit Schädelhirntraumata stützt, sieht eine Störung in den Exekutivfunktionen als Ursache für kommunikativ-kognitive Defizite (Martin & McDonald, 2003). Die vorliegende Arbeit folgt insbesondere der letztgenannte Theorie, da die Diagnose kommunikativ-kognitiver Defizite anhand der Evaluation von Leistungen bei der Wortgenerierung erfolgen soll. Diese erfordern, wie beschrieben, für eine erfolgreiche Lösung in einem besonderen Maße auch intakte kognitiv-exekutive Funktionen.

Wortgenerierungsleistungen bei kommunikativ-kognitiven Defiziten

Es existieren zahlreiche Studien, die sich mit der Analyse von Wortgenerierungsleistungen bei Schädelhirntraumata bzw. Frontalhirnläsionen befassen (vgl. Bittner & Crowe, 2006; Capitani, Rosci, Saetti, & Laiacona, 2009; Raskin & Rearick, 1996) und auch Testverfahren wie die *Scales of Cognitive Ability for Traumatic Brain Injury* (SCATBI; Adamovich & Henderson, 1992) überprüfen neben anderen kognitiven und linguistischen Funktionen die Wortgenerierungsleistungen bei dieser klinischen Population.

Reduzierte Leistungen werden in lexikalischen Wortgenerierungsaufgaben (Bittner & Crowe, 2006; Kavé et al., 2011), aber auch für die semantische Wortgenerierungsleistung berichtet (Kavé et al., 2011). Kontroverse Ergebnisse bestehen bezüglich der Frage, ob bei Patienten mit frontalen Läsionen eine Dominanz beeinträchtigter lexikalischer Wortgenerierungsleistungen zu finden ist oder semantische und lexikalische Wortgenerierungsleistungen gleichermaßen defizitär sind (Capitani et al., 2009). Capitani und Kollegen (2009) fanden für die Gruppe der Schädelhirntraumapatienten schlechtere Leistungen in der lexikalischen im Vergleich zur semantischen Wortgenerierung (siehe auch Jurado, Mataro, Verger, Bartumeus, & Junque, 2000). Die Autoren betonen jedoch auch, dass einige Patienten pathologische Leistungen in beiden Aufgaben zeigten. Henry und Crawford (2004) konnten zeigen, dass semantische und lexikalische Wortgenerierungsleistungen ein ähnlich sensitives Indiz für Frontalhirnläsionen sind. Dieses bestätigen auch Baldo et al. (2001): semantische und lexikalische Wortgenerierungsleistungen zeigten sich bei Patienten mit Frontalhirnläsionen gleichermaßen beeinträchtigt, wobei insgesamt schlechtere Leistungen für Patienten mit links-frontaler im Vergleich zu rechts-frontaler Läsion gefunden werden konnten. Spezifische Aussagen über den Zusammenhang kommunikativ-kognitiver Defizite und Wortgenerierungsleistungen sind nur bedingt zu treffen. Viele Studien untersuchen zwar Wortgenerierungsleistungen in klinischen Populationen für die kommunikativ-kognitive Defizite häufig berichtet werden, gehen aber

nicht explizit darauf ein, ob bei der untersuchten Stichprobe tatsächlich kommunikativ-kognitive Defizite bestanden.

2.4.3 Beeinträchtigungen von Sprache und Kommunikation bei Multipler Sklerose

In den beiden vorausgehenden Abschnitten wurden sprachsystematische und kommunikativ-kognitive Defizite verschiedener Ätiologien dargestellt. An dieser Stelle folgt nun die Betrachtung von Sprach- und Kommunikationsdefiziten bei einem spezifischen neurologischen Störungsbild, der Multiplen Sklerose. Während in den Abschnitten 2.4.1 und 2.4.2 die Beschreibung der Störungsqualität im Vordergrund stand, folgt nun mehr oder weniger eine Anwendung der beschriebenen Defizite auf Sprach- und Kommunikationsstörungen bei Multipler Sklerose. Diese Struktur wurde gewählt, da für Multiple Sklerose noch nicht endgültig geklärt ist, ob die beobachtbaren Sprach- und Kommunikationsdefizite sekundärer, das heißt kommunikativ-kognitiver, oder aber sprachsystematischer Natur sind.

Die Multiple Sklerose ist eine der häufigsten neurologischen Erkrankungen des jungen und mittleren Erwachsenenalters, deren Ätiologie bis heute nicht hinreichend geklärt ist (Poeck & Hacke, 2001; Pusswald & Vass, 2011). Sie wird als entzündliche Erkrankung des zentralen Nervensystems beschrieben, welche vor allem durch eine fokale Demyelinisierung in Bereichen mit hoher Dichte an weißer Substanz gekennzeichnet ist und dadurch häufig auch als Entmarkungserkrankung bezeichnet. Neben fokalen Demyelinisierungsherden sind jedoch auch und vor allem mit Fortschreiten der Erkrankung axonale Schädigungen zu finden.¹⁴ Als Prädilektionsstellen der Demyelinisierungsherde werden unter anderem das periventrikuläre Marklager, das Marklager des Kleinhirns, die Sehnerven, das Corpus Callosum, der Hirnstamm oder die Pons beschrieben (Poeck & Hacke, 2001; Pusswald & Vass, 2011). Aber auch eine Beteiligung von Läsionen in der grauen Substanz wird in der Literatur immer weiter hervorgehoben¹⁵, insbesondere bei Progression der Erkrankung. Myelinisierte Axone sind auch in der grauen Substanz des zentralen Nervensystems zu finden, daher können Demyelinisierungsherde auch in Arealen mit grauer Substanz, sowohl kortikal als auch subkortikal auftreten. Genannt werden diesbezüglich unter anderem der Motorkortex sowie der frontale, temporale und parietale Kortex, der Gyrus Cinguli, das Kleinhirn, der Thalamus, die Basalganglien, die Substantia Nigra, die Amygdala

¹⁴Weitere Angaben zur Pathophysiologie der Multiplen Sklerose finden sich z.B. in Pusswald & Vass (2011).

¹⁵Demyelinisierungsherde in der grauen Substanz werden schon seit Beginn des 20. Jahrhunderts berichtet, in Beschreibungen der Multiplen Sklerose jedoch häufig vernachlässigt. Demyelinisierungen in der grauen Substanz sind mit herkömmlichen Magnet-Resonanz-Tomographie-Techniken nicht nachweisbar. Immunhistochemische Analysen (post-mortem) erlauben bessere Rückschlüsse (Geurts, Calabrese, Fisher, & Rudick, 2012).

sowie die graue Substanz des Rückenmarks (Geurts et al., 2012). Insgesamt lässt sich festhalten, dass die Lokalisationen und ihre mögliche Verteilung im Gehirn sehr heterogen sind und sich bei keinen zwei Patienten dieselben Läsionsherde finden lassen (Pusswald & Vass, 2011). Ähnlich heterogen sind daher auch die resultierenden Beeinträchtigungen.

Es werden verschiedene Verlaufsformen der Multiplen Sklerose unterschieden (Poeck & Hacke, 2001; Pusswald & Vass, 2011); Abbildung 2.7 veranschaulicht die Wichtigsten. Bei etwa 85 bis 90% der Erkrankten beginnt die Multiple Sklerose schubförmig, das heißt, es kommt akut zum Auftreten einer neurologischen Symptomatik (sogenannter *Schub*), die sich innerhalb weniger Tage bis Wochen vollständig oder auch nur teilweise zurückbildet (schubförmig-remittierende Multiple Sklerose)¹⁶. Bei 50 bis 80% der Erkrankten geht diese initial schubförmig-remittierend verlaufende Multiple Sklerose nach etwa 10 Jahren in eine progrediente Verlaufsform über. Es kommt zu einer kontinuierlichen Verschlechterung der neurologischen Symptomatik (sekundär chronisch-progrediente Multiple Sklerose). Ein primär chronisch-progredienter Verlauf liegt bei etwa 10% der Betroffenen vor (Chiaravalloti & Luca, 2008; Poeck & Hacke, 2001; Pusswald & Vass, 2011). Die Unterscheidung der verschiedenen Verlaufsformen ist insofern relevant, als mit ihr einhergehend unterschiedliche Ausprägungen und Schweregrade funktioneller sowie kognitiver Defizite diskutiert werden.

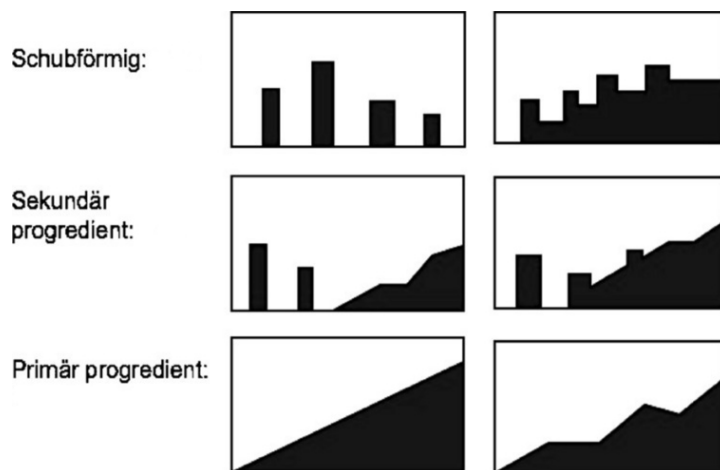


Abbildung 2.7: Darstellung der Verlaufsformen bei Multipler Sklerose (modifiziert); aus Pusswald & Vass (2011, S. 332)

¹⁶Die initiale Reversibilität der Symptome lässt sich dadurch erklären, dass zerstörtes Myelin durch Oligodendrozyten wieder nachgebildet werden kann (Pusswald & Vass, 2011).

Klinische Symptome

Bedingt durch die große Variabilität der möglichen Lokalisationen von Demyelinisierungsherden sind auch die im Rahmen einer Multiplen Sklerose auftretenden Symptome sehr heterogen (Poeck & Hacke, 2001; Pusswald & Vass, 2011). Häufig auftretende Symptome betreffen den Visus (Optikusneuritis bzw. retrobulbäre Optikusneuritis) und die Okulomotorik, zeigen sich in Funktionsstörungen des Nervus Facialis und des sensiblen Nervus Trigemini (Trigeminusneuralgie) oder in Form zentraler Paresen. Als weitere frequente Symptome gelten Sensibilitätsstörungen, Inkontinenz, Charcot-Trias (Nystagmus, Intentionstremor, skandierendes Sprechen) als Ausdruck von Kleinhirnfunktionsstörungen, psychische Veränderungen und Fatigue (Poeck & Hacke, 2001; Pusswald & Vass, 2011). Für eine detaillierte Übersicht zu den Symptomen bei Multipler Sklerose siehe beispielsweise Poeck (2001) oder Pusswald & Vass (2011). Sprachtherapeutisch werden Patienten mit Multipler Sklerose vor allem bezüglich dysarthrischer und dysphagischer Defizite behandelt. Insbesondere dysarthrische Störungen sind auch bei der Evaluation von Wortgenerierungsleistungen von Bedeutung, da die im Rahmen einer Dysarthrie mögliche Verlangsamung des Sprechtempos sekundär eine Reduzierung der Wortgenerierungsleistung bedingen kann (Chiaravalloti & Luca, 2008). Es gilt daher, bei gleichzeitigem Bestehen einer motorischen bedingten Verlangsamung des Sprechtempos, quantitativ reduzierte Leistungen bei der Wortgenerierung vorsichtig zu interpretieren, um einer Überinterpretation vorzubeugen.

Von besonderer Bedeutung sind darüber hinaus kognitive Defizite, welche bereits in frühen Stadien der Erkrankung auftreten und sich negativ auf die Lebensqualität sowie die Partizipation auswirken können (Chiaravalloti & Luca, 2008; Rao, Leo, Ellington, Nauertz, Bernardin, & Unverzagt, 1991b; Zakzanis, 2000). Beeinträchtigungen kognitiver Funktionen werden bei etwa der Hälfte der an Multipler Sklerose erkrankten Patienten beschrieben (Pusswald & Vass, 2011; Rao, Grafman, DiGiulio, Mittenberg, Bernardin, Leo, Luchetta, & Unverzagt, 1991a). Die Ausprägung der Defizite variiert dabei interindividuell, wobei vor allem höhere kognitive Funktionen betroffen sein sollen (Chiaravalloti & Luca, 2008). Kognitive Defizite betreffen häufig Bereiche des Gedächtnisses, das visuelle Lernen oder die Informationsverarbeitungsgeschwindigkeit (Chiaravalloti & Luca, 2008; Pusswald & Vass, 2011; Rao et al., 1991a). Aber auch komplexe Aufmerksamkeitsprozesse und Exekutivfunktionen gelten als beeinträchtigt (Chiaravalloti & Luca, 2008; Pusswald & Vass, 2011). Gedächtnisdefizite werden vornehmlich in Bezug auf Arbeits- und Langzeitgedächtnisleistungen beschrieben (Chiaravalloti & Luca, 2008; Zakzanis, 2000), wobei die visuell-räumliche Verarbeitung im Vergleich zur verbalen als stärker beeinträchtigt gilt (Pusswald & Vass, 2011). Verlangsamte Reaktionszeiten von Patienten mit Multipler

Sklerose werden als Hinweis auf eine reduzierte Verarbeitungsgeschwindigkeit gesehen (Rao, St. Aubin-Faubert, & Leo, 1989; Zakzanis, 2000). Rao und Kollegen (1989) gehen sogar soweit und sprechen von einer generellen kognitiven Verlangsamung bei Multipler Sklerose. Während basale Aufmerksamkeitsleistungen als weitestgehend intakt gelten, sind Patienten mit Multipler Sklerose bei solchen Aufgaben beeinträchtigt, die die Vigilanz und insbesondere die geteilte Aufmerksamkeit erfordern (Chiaravalloti & Luca, 2008). Defizite in Aufmerksamkeitsprozessen gelten als erste Hinweise auf kognitive Defizite bei Multipler Sklerose (Pusswald & Vass, 2011).

Beeinträchtigungen in den Exekutivfunktionen sind typisch und äußern sich in einer Minderung der kognitiven Flexibilität, bei der zielgerichteten Handlungsplanung oder der Konzeptbildung. Als besonders sensitiv für die Identifizierung kognitiver Defizite bei Multipler Sklerose werden Wortgenerierungsleistungen beschrieben (Beatty, 2002; Henry & Beatty, 2006; Miller & Leary, 2007). Darüber hinaus sollen Perseverationstendenzen als Hinweis auf Defizite in den Exekutivfunktionen zu beobachten sein (Butters, Goldstein, Allen, & Shemansky, 1998; Chiaravalloti & Luca, 2008; Pusswald & Vass, 2011). Es wird kontrovers diskutiert, ob das kognitive Profil von Patienten mit Multipler Sklerose dem einer subkortikalen Demenz entspricht (vgl. hierzu Rao, 1986, siehe demgegenüber aber auch z.B. Butters et al., 1998). Nach Drake, Carrá, Allegri & Luetic (2006) sollen kognitive Defizite bei schubförmig-remittierendem Verlauf einem eher subkortikalen Muster entsprechen, während bei chronisch-progredientem Verlauf umfassendere, weitergehende Defizite gefunden werden.

Der Zusammenhang von Verlaufsformen der Multiplen Sklerose und der Manifestation kognitiver Dysfunktionen ist jedoch nicht abschließend geklärt (Nocentini, Rossini, Carlesimo, Graceffa, Grasso, Lupoi, Oliveri, Orlacchio, Pozzilli, Rizzato, & Caltagirone, 2001). Während es Studien gibt, die einen kausalen Zusammenhang zwischen der Verlaufsform der Multiplen Sklerose und kognitiven Defiziten sehen und ausgeprägtere und qualitativ unterschiedliche Beeinträchtigungen sowie eine frühere Manifestation bei chronisch-progredientem Verlauf beschreiben (Drake et al., 2006; Huijbregts, Kalkers, de Sonneville, de Groot, Reuling & Polman, 2004; Kraus, Schütze, Brokate, Kröger & Schwendemann, 2005; Nocentini et al., 2001; Ruet, Deloire, Charré-Morin, Hamel & Brochet, 2013; Zakzanis, 2000), finden beispielsweise Wishart und Sharpe (1997) keinen Unterschied zwischen Patienten mit schubförmig-remittierender und chronisch-progredienter Multipler Sklerose bezüglich der Schwere kognitiver Beeinträchtigungen. Es gibt Hinweise darauf, dass insbesondere Defizite in den Exekutivfunktionen bei Patienten mit chronisch-progredientem Verlauf im Vergleich zu Patienten mit schubförmig-remittierendem Verlauf stärker ausgeprägt sind. Dies wird als Indiz dafür gewertet, dass mit Progression der Erkrankung zunehmend Demyelinisierungsherde in frontalen Hirnarealen zu finden sind

(Kraus et al., 2005; Zakzanis, 2000). Ob es sich hierbei um tatsächliche Unterschiede handelt oder nicht vielmehr Unterschiede in Alter, Krankheitsdauer, Schwere der neurologischen Defizite oder eine Verlangsamung des Sprechtempos ursächlich sind, wird kontrovers diskutiert (Henry & Beatty, 2006). Hinweise darauf, dass es sich um tatsächliche Unterschiede handeln könnte, lieferten kürzlich Ruet und Kollegen (2013). Die Autoren untersuchten in einer aktuellen Studie die kognitiven Fähigkeiten von Patienten mit chronisch-progredienter und schubförmig-remittierender Multipler Sklerose sowie Sprachgesunder in unterschiedlichen kognitiven Domänen, wobei sie die Faktoren Alter, Geschlecht, Bildungsgrad und Schweregrad der Erkrankung kontrollierten. Die Autoren zeigten, dass Patienten mit chronisch-progredienter Multipler Sklerose im Vergleich zu Patienten mit schubförmigem Verlauf der Erkrankung sowohl stärker als auch umfassender (d.h. in mehr Domänen) beeinträchtigt waren. Diese Unterschiede zeigten sich auch dann noch, wenn der Faktor Schweregrad der Erkrankung kontrolliert wurde (Ruet et al., 2013).

Uneinigkeit herrscht auch bezüglich der Manifestation sprachlicher Beeinträchtigungen im Rahmen einer Multiplen Sklerose (Chiaravalloti & Luca, 2008; Pusswald & Vass, 2011; Zakzanis, 2000). In wenigen Fällen werden Symptome aphasischer Sprachstörungen beschrieben (Lacour et al., 2004). Häufiger wird von auftretenden Wortfindungsstörungen, reduzierten Wortgenerierungsleistungen oder *high-level-language* Defiziten berichtet (Kujala, Portin, & Ruutiainen, 1996; Laakso, Brunnegård, Hartelius, & Ahlsén, 2000; Lethlean & Murdoch, 1997). Dabei werden unter anderem Defizite beim Verstehen von Metaphern und ambigen Sätzen oder beim Schlussfolgern beschrieben (Laakso et al., 2000; Lethlean & Murdoch, 1997). Auch bei Aufgaben zu Wortschatz und Semantik (z.B. beim Assoziieren oder Finden von Antonymen) zeigen Patienten mit Multipler Sklerose schlechtere Leistungen als Sprachgesunde (Lethlean & Murdoch, 1997). Diese Defizite werden sowohl für Patienten mit chronisch-progredienter als auch für Patienten mit schubförmig-remittierender Multipler Sklerose beschrieben, wobei sich die *high-level-language*-Defizite in vielen Aufgaben in einem stärkeren Ausmaß bei Patienten mit chronisch-progredientem Verlauf manifestieren sollen (Lethlean & Murdoch, 1997). Lethlean und Murdoch (1994) berichten insbesondere von semantischen Fehlern in einer Aufgabe zum konfrontativen Benennen und konnten diese auf eine Störung im Zugriff auf das semantische System zurückführen. Auch Drake und Kollegen (2006) beschreiben semantische Defizite und betonen dabei vor allem eine defizitäre Organisation im semantischen System bzw. semantischer Netzwerke sowie Schwierigkeiten beim Enkodieren gespeicherter Information. Aber auch Defizite beim lexikalischen Zugriff werden geschildert und als Ausdruck einer verminderten kognitiven Flexibilität mit *high-level language* Defiziten in Verbindung gebracht (Sepulcre et al., 2011).

Sprachliche Beeinträchtigungen werden auf den verschiedensten Ebenen, so auch in der Spontansprache bzw. beim narrativen Erzählen berichtet (Arnott, Jordan, Murdoch, & Lethlean, 1997; Arrondo, Sepulcre, Duque, Toledo, & Villoslada, 2010; Böwering, 2012). Unterschiede zu Sprachgesunden zeigen sich dabei in Quantität und Qualität der Äußerungen. So werden für Multiple Sklerose Patienten beispielsweise eine weniger kohärente Sprache, ein größerer Anteil von Wortfindungsstörungen oder syntaktisch weniger komplexe bzw. kürzere Sätze genannt (Arnott et al., 1997; Arrondo et al., 2010; Böwering, 2012).

Insgesamt sind die auftretenden Defizite häufig subtil und es mangelt an geeigneten Verfahren, sie aufzudecken und angemessen zu interpretieren (Laakso et al., 2000). Die im Rahmen einer Multiplen Sklerose zu beobachtenden Defizite entsprechen weitestgehend den bereits in Abschnitt 2.4.2 als kommunikativ-kognitive Defizite beschriebenen Symptomen und sind daher in vielen Fällen auch als solche zu klassifizieren. Aufgrund der großen Variabilität möglicher Läsionslokalisationen sind aber auch sprachsystematische Störungen denkbar. Eine sichere Möglichkeit der Abgrenzung gibt es diesbezüglich bislang nicht.

Wortgenerierungsleistungen bei Multipler Sklerose

Als eines der sensitivsten Maße kognitiver Beeinträchtigungen bei Multipler Sklerose wird die Evaluation von Wortgenerierungsleistungen beschrieben (Beatty, 2002; Henry & Beatty, 2006; Miller & Leary, 2007). Wortgenerierungsaufgaben sind daher auch Bestandteil verschiedener Testverfahren zur Identifizierung kognitiver Defizite bei Multipler Sklerose. So überprüft das *Multiple Sklerose Inventar Cognition* (MUSIC; Calabrese, Kalbe & Kessler, 2004) beispielsweise die Wortgenerierungskompetenz in einer intradimensionalen semantischen Wechsel-Aufgabe (Tiere/Möbel) oder die *Neuropsychological Screening Battery for Multiple Sclerosis* (Rao et al., 1990) lexikalische Wortgenerierungsleistungen.

Kontroverse Ergebnisse (ähnlich denen bei kommunikativ-kognitiven Defiziten) bestehen jedoch bezüglich der Frage nach Leistungsunterschieden in den verschiedenen Aufgaben. Von vergleichbaren Leistungen in semantischen und lexikalischen Wortgenerierungsaufgaben (Beatty, Goodkin, Beatty & Monson, 1989; Beatty, 2002; Kraus et al., 2005; Messinis et al., 2013) wird ebenso berichtet wie von größeren Beeinträchtigungen in semantischen (Sepulcre et al., 2011; Zakzanis, 2000) oder lexikalischen (Henry & Beatty, 2006; Nocentini et al., 2001) Wortgenerierungsleistungen. Es ist also noch nicht abschließend geklärt, ob den eingeschränkten Wortgenerierungsleistungen bei Multipler Sklerose ein primär semantisches oder kognitiv-exekutives

Defizit zugrunde liegt.

Beatty und Kollegen (2002) wie auch Henry & Beatty (2006) beschreiben insgesamt deutliche Beeinträchtigungen bei der Wortgenerierung und betonen, dass semantische und lexikalische Wortgenerierungsleistungen ähnlich gut geeignet sind, um zwischen Patienten mit Multipler Sklerose und neurologisch gesunden Kontrollprobanden zu dissoziieren. Demgegenüber berichten Zakzanis et al. (2000), dass Leistungen in semantischen Wortgenerierungsaufgaben besser zwischen Patienten mit Multipler Sklerose und Sprachgesunden unterscheiden als Leistungen in lexikalischen Wortgenerierungsaufgaben. Eine verlässliche Dissoziation war jedoch nur für etwa die Hälfte der Patienten möglich und dabei reliabler für Patienten mit chronisch-progredientem im Vergleich zu schubförmig-remittierendem Verlauf. Dies könnte auf eine Progredienz in Richtung eines zunehmenden semantischen Defizits hinweisen.

Aber auch das Auftreten von Unterschieden in Wortgenerierungsleistungen bei Patienten mit unterschiedlichen Verlaufsformen einer Multiplen Sklerose wird noch kontrovers diskutiert. Studien weisen darauf hin, dass vor allem Leistungen in der Wortgenerierung geeignet sein sollen um verschiedene Verlaufsformen einer Multiplen Sklerose voneinander abzugrenzen (Kraus et al., 2005; Miller & Leary, 2007 aber vgl. auch Huijbregts et al., 2004). Kraus und Kollegen (2005) fanden, dass insbesondere Leistungen in semantischen Wortgenerierungsaufgaben ein geeignetes Maß zu sein scheinen, um Patienten mit primär und sekundär chronisch-progredienter Multipler Sklerose voneinander zu dissoziieren. Patienten mit primär chronisch-progredienter Multipler Sklerose produzierten im Vergleich zu Patienten mit sekundär chronisch-progredientem Verlauf der Erkrankung weniger korrekte Nennungen in semantischen Wortgenerierungsaufgaben, jedoch nicht in lexikalischen. Huijbregts und Kollegen (2004) zeigten, dass Patienten mit schubförmig-remittierender Multipler Sklerose, obwohl sie generell bessere kognitive Leistungen im Vergleich zu Patienten mit chronisch-progredientem Verlauf der Multiplen Sklerose zeigten, dennoch relativ geringe Leistungen in Wortgenerierungsaufgaben aufwiesen. Ein chronisch-progredienter Verlauf soll darüber hinaus im Vergleich zur schubförmig-remittierenden Variante einer Multiplen Sklerose mit größeren Beeinträchtigungen in semantischen und lexikalischen Wortgenerierungsleistungen einhergehen (Huijbregts et al., 2004). Inwieweit das einem tatsächlichen Unterschied entspricht oder auf Unterschiede in Alter, Krankheitsdauer, Schwere der neurologischen Defizite, oder ein sprechmotorisches Defizit zurückzuführen ist, ist auch hier bislang nicht geklärt (Henry & Beatty, 2006). Messinis et al. (2013) fanden diesen Unterschied nicht. Patienten mit schubförmig-remittierender und chronisch-progredienter Multipler Sklerose unterschieden sich in der Studie der Autoren nur bezüglich der strategischen *Switches* bei der lexikalischen Wortgenerierung. Es bestehen demnach noch Kontroversen bezüglich eines

möglichen Zusammenhangs zwischen der Verlaufsform einer Multiplen Sklerose und dem Ausmaß der Leistungsbeeinträchtigung bei der Wortgenerierung.

Der nun folgende Teil widmet sich diagnostischen Aspekten, die bei der Evaluation der Leistungen zur diagnostischen Abklärung von Sprach- und Kommunikationsstörungen von Relevanz sind.

2.5 Diagnostik

Eine eingehende Befundung stellt den Beginn einer jeden sprachtherapeutischen Intervention dar und beruht im Wesentlichen auf der Beobachtung von Verhalten (Lezak et al., 2004). Sie beinhaltet, bezogen auf eine neurolinguistische oder -psychologische Diagnostik, zum einen die Beobachtung von Leistungen in spezifischen kognitiven, linguistischen oder kommunikativen Aufgaben, wie z.B. beim Bildbenennen, in Aufgaben zur Überprüfung von Arbeitsgedächtnisleistungen oder bei der Wortgenerierung. Zum anderen bedarf sie aber auch ganz allgemein der Beobachtung des Verhaltens bei der Ausführung einer solchen Aufgabe oder in der alltäglichen Interaktion und Kommunikation (Lezak et al., 2004). Zur Identifizierung spezifischer Beeinträchtigung sind dann präzise und sensitive Leistungsindikatoren erforderlich (Lezak et al., 2004). Das heißt, es müssen verlässliche Hinweise für das Vorliegen eines Defizits zu erkennen sein, die idealerweise eine reliable Identifizierung der Beeinträchtigung ermöglichen. Auch ein spezifisches Diagnostikum sollte sich daher durch eine hohe *Sensitivität* und *Spezifität* auszeichnen (Keil & Kaszniak, 2002). Die *Sensitivität* gibt an, mit welcher Wahrscheinlichkeit ein Test positiv ausfällt, wenn man tatsächlich krank ist, das heißt, wie sicher kann ein Test die Erkrankung identifizieren. Dies ist vor allem bei subtilen Beeinträchtigungen von Bedeutung. Die *Spezifität* bezeichnet hingegen die Wahrscheinlichkeit, mit der ein Test negativ ausfällt, wenn keine (oder eine andere) Störung vorliegt (vgl. auch Bortz & Lienert, 2008).¹⁷

Eine Diagnostik, hinsichtlich neurolinguistischer und/oder -psychologischer Evaluation, zielt in erster Linie auf eine sichere Identifizierung von pathologischen Leistungen und damit einhergehend auf eine Abgrenzung von nicht-pathologischen, normal-sprachlichen Leistungen ab (Wehmeyer & Grötzbach, 2006). Darüber hinaus ist eine genaue Identifizierung der Symptome für eine spezifische Beschreibung des zugrunde liegenden Leistungsde-

¹⁷Sensitivität und Spezifität dürfen nicht mit dem positiven bzw. negativen Vorhersagewert verwechselt werden, welcher angibt, mit welcher Wahrscheinlichkeit eine Störung vorliegt bzw. nicht vorliegt, wenn der Test positiv bzw. negativ ausfällt, siehe hierzu Bortz & Lienert (2008).

fizits erforderlich. Kaplan (1988) betont dies in der Darstellung ihres prozessorientierten Ansatzes. Für die Einschätzung und Charakterisierung eines Leistungsdefizits ist demnach nicht nur das in einem spezifischen Test erzielte Ergebnis von Bedeutung. Auch die einzelnen Schritte, die zu diesem Ergebnis geführt haben, sollten für die Einschätzung der Leistung herangezogen werden, da ein und derselbe Testwert oftmals durch qualitativ unterschiedliche Leistungen erreicht werden kann (Kaplan, 1988). Das Ergebnis eines spezifischen Tests (ein Test-Score) entsteht nach Lezak et al. (2004) durch indirekte Beobachtung und stellt damit eine Zusammenfassung der Leistung dar, welche sich in quantitativen, numerischen Daten ausdrückt. Nuancen einer Störung werden häufig nur durch eine qualitativ-deskriptive Herangehensweise offensichtlich (Lezak et al., 2004). Während quantitative Aspekte einer Leistungen besser meßbar und objektivierbar sind, kann eine qualitative Analyse zu einer spezifischeren Beschreibung eines Defizits beitragen (Lezak et al., 2004). Eine gute Diagnostik sollte daher sowohl quantitative als auch qualitative Aspekte berücksichtigen (Lezak et al., 2004). Übertragen auf die Evaluation von Wortgenerierungsleistungen bedeutet dies, dass neben der *Anzahl korrekter Nennungen* als quantitatives Leistungsmaß, eine Untersuchung der einzelnen Nennungen und ihrer spezifischen Charakteristika qualitative Aspekte zur Einschätzung des Leistungsdefizits beitragen kann. Generell kann eine qualitative, deskriptive Analyse sprachlicher Äußerungen (ob auf Einzelwortebene bei der Wortgenerierung oder in der Spontansprache) dazu beitragen, die einem Defizit zugrunde liegenden Prozesse aufzudecken (Abwender, Swan, Bowerman & Connely, 2001; Lezak et al., 2004). Auf Basis dieser Erkenntnisse sind dann eine fundierte Therapieplanung und ein besseres Monitoring der identifizierten Störung über den Therapieverlauf hinweg möglich (Kaplan, 1988).

Die Erfassung quantitativer sowie qualitativer Leistungen bedarf verschiedener Herangehensweisen und Verfahren (Lezak et al., 2004). Im Wesentlichen wird zwischen *standardisierten* und *nicht-standardisierten* Diagnostikverfahren¹⁸ differenziert (Coelho et al., 2005; Tesak, 2006; Turkstra et al., 2005a,b). Als standardisierte Testverfahren gelten solche, bei denen auf der Basis spezifischer Testitems vorgegebene und explizit definierte Anweisungen zur Durchführung existieren. Nach Turkstra und Kollegen (2005a) besteht ein standardisiertes Vorgehen aus *Evaluation* (d.h. Untersuchung zur Beurteilung von Wertigkeit, Qualität, Wichtigkeit, Ausmaß oder Zustand) und *Assessment* (d.h. eine Beurteilung basierend auf dem Verständnis der jeweiligen Situation). Dabei stellt die *Evaluation* den Prozess der Untersuchung dar, während das *Assessment* das Ergebnis der Untersuchung

¹⁸Weitere mögliche Unterscheidungskriterien sind beispielsweise formal vs. informell, normiert vs. kriterienbezogen, statisch/deskriptiv vs. dynamisch/experimentell, initial vs. outcome, vgl. hierzu auch Coelho et al. (2005)

fokussiert (Turkstra et al., 2005a). Eine weitere wichtige Komponente standardisierter Verfahren ist die der Normierung. Normierte Verfahren ermöglichen den Vergleich einer individuellen Leistung mit einer Referenzstichprobe, beispielsweise anhand von Prozenträngen (Turkstra et al., 2005b; Coelho et al., 2005). Nicht-standardisierte Verfahren sind demgegenüber solche, für die keine vorgegebenen Prozeduren und Items bestehen. Häufig handelt es sich um individuell zusammengestellte Aufgaben zur Einschätzung spezifischer Leistungen bzw. deren Beeinträchtigung (Coelho et al., 2005; Tesak, 2006). Als besonders geeignet gelten nicht-standardisierte Verfahren für die Einschätzung von Leistungen, für die noch keine standardisierten Verfahren existieren sowie für die Überprüfung verschiedener Alltagskompetenzen und der Kommunikation im Alltag sowie für die systematische Beurteilung qualitativer Leistungen (Coelho et al., 2005). Auch für Patienten mit vergleichsweise subtilen kommunikativ-kognitiven Beeinträchtigungen wird die Relevanz nicht-standardisierter Verfahren bei der Diagnosestellung betont, da die Beeinträchtigungen bei Patienten dieser Population insbesondere durch unauffällige Leistungen in strukturierten Verfahren und auffälliger Performanz in weniger strukturierten Settings charakterisiert sind (Coelho et al., 2005).

Das Ziel einer fundierten Diagnostik besteht demnach nicht darin, den identifizierten Symptomen und Symptomkombinationen einen Namen zu geben, das heißt es als Aphasie, kommunikativ-kognitives Defizit oder ähnlich bezeichnen zu können. Aufgabe einer Diagnostik ist vielmehr eine aussagekräftige Beschreibung und Identifizierung zugrunde liegender Leistungsdefizite, um davon ausgehend eine auf den Patienten und seine individuelle Beeinträchtigung ausgerichtete Behandlung planen zu können (Wehmeyer & Grötzbach, 2006). Die Verwendung spezifischer Störungs- und/oder Syndrom-Bezeichnungen (wie im Falle der Aphasie) nach Abschluss der Diagnostik kann und sollte lediglich als Möglichkeit einer zusammenfassenden Beschreibung verstanden werden, die eine Kommunikation zwischen verschiedenen Disziplinen erleichtert und dadurch eine zügige Einleitung therapeutischer Interventionen faszilitieren kann (Tesak, 2006). Insgesamt sollte die individuelle Erfassung und Beschreibung einer Leistungen und ihrer Störungen, auch vor dem sozialen und beruflichen Hintergrund des Patienten, stets im Vordergrund stehen (Tesak, 2006).

Während für die Diagnose sprachsystematischer Störungen wie der Aphasie zahlreiche standardisierte und normierte Testverfahren existieren, die eine reliable Identifizierung der Störung ermöglichen¹⁹ (z.B. der *Aachener Aphasia Test* von Huber, Poeck, Weniger & Willmes, 1983; die *Boston Dia-*

¹⁹Eine Ausnahme stellt hier sicherlich die Diagnose von Restaphasien dar, da standardisierte Testverfahren in der Regel nicht weiter zwischen *Restaphasie* vs. *keine Aphasie* differenzieren; vgl. hierzu z.B. Jaecks, Hielscher-Fastabend & Stenneken (2012) oder Huber et al. (1983).

gnostic Aphasia Examination von Goodglass & Kaplan, 1972; oder die *Western Aphasia Battery* von Kertesz, 1982), ist dies für die Diagnose kommunikativ-kognitiver Defizite oder bei Sprach- und Kommunikationsstörungen im Rahmen progredienter neurogener Erkrankungen wie beispielsweise der Multiplen Sklerose bislang nicht gegeben (Coelho et al., 2005; Duff & Proctor 2002; Turkstra et al., 2005a; Turkstra et al., 2005b; Vogel, Maruff & Morgan, 2010). Die Diagnose kommunikativ-kognitiver Defizite und ihre Abgrenzung von sprachgesunden Leistungen stellt vielmehr immer noch eine Herausforderung dar (Turkstra et al., 2005b). Existierende Verfahren betrachten vor allem quantitative Aspekte und fokussieren häufig die Überprüfung kognitiver Funktionen, wie beispielsweise das *Multiple Sklerose Inventar Cognition* (MUSIC; Calabrese et al., 2004) oder die *Scales of Cognitive Ability for Traumatic Brain Injury*²⁰ (SCATBI; Adamovich & Henderson, 1992). Darüber hinaus werden im englischen Sprachraum Verfahren zur Diagnose von Beeinträchtigungen höherer sprachlicher Funktionen wie z.B. der *Test of Language Competence-Expanded Edition* (TLC-Expanded; Wiig & Secord, 1989) oder *The WORD Test-2: Adolescent* (Bowers, Huisingh, LoGiudice, & Orman, 2005) eingesetzt, welche jedoch vornehmlich zur Überprüfung sprachlicher Kompetenzen während der Entwicklung bis ins junge Erwachsenenalter konzipiert und normiert wurden. Eine reliable Erfassung subtiler kommunikativ-kognitiver Defizite im Erwachsenenalter ist damit nicht möglich. Für den deutschsprachigen Raum besteht insgesamt ein Mangel an geeigneten Prozeduren und Standards zur Überprüfung von pragmatischen Funktionen und höheren sprachlichen Kompetenzen sowie auch an Verfahren zur Erfassung von Leistungen in den Exekutivfunktionen (Drechsler, 2007).

Um die Herausforderung einer reliablen Diagnosestellung zu meistern, ist es wichtig, sich auf die Mechanismen, Prozesse oder Funktionen zu fokussieren, die einer kommunikativ-kognitiven Beeinträchtigung (wahrscheinlich) zugrunde liegen (Martin & McDonald, 2003). Dazu braucht eine gute Diagnostik reliable und interpretierbare Maße kognitiver Funktionen (Keil & Kaszniak, 2002). Wie in Abschnitt 2.4.2 beschrieben, werden kommunikativ-kognitive Defizite im Zusammenhang mit einer Störung der Exekutivfunktionen gesehen. Ein Ansatz ist daher beispielsweise, sich der Diagnose einer kommunikativ-kognitiven Beeinträchtigung über die Evaluation der Leistungen in den Exekutivfunktionen zu nähern.

Für die Identifizierung von Exekutivfunktionsdefiziten gelten insbesondere solche Aufgaben als geeignet, die nicht auf Basis etablierter Routinen

²⁰Die SCATBI (Adamovich & Henderson, 1992) untersucht darüber hinaus auch linguistische Aspekte und wird von den Autoren als Test zur Überprüfung kognitiv-linguistischer Funktionen nach Schädelhirntrauma vorgeschlagen (vgl. auch Adamovich & Hutchinson, 1994). Der Schwerpunkt des Verfahrens liegt jedoch auf der Überprüfung kognitiver Fähigkeiten.

ablaufen, das heißt Aufgaben, die neue Anforderungen an eine Person stellen oder deren Lösung ein Abweichen von etablierten Routinen erfordert (Keil & Kaszniak, 2002). Diese Forderungen erfüllen Aufgaben zur Elizitierung von Wortgenerierungsleistungen. Insbesondere lexikalische Wortgenerierungsaufgaben gelten als geeignet, da diese für eine angemessene Lösung der Aufgabe, ein Abweichen von etablierten semantischen Suchstrategien erfordern (Basso et al., 1997; Keil & Kaszniak, 2002). Wie beschrieben (vgl. Abschnitt 2.3) beruhen Wortgenerierungsleistungen nicht nur auf intakten kognitiv-exekutiven Funktionen wie *Initiierung*, *Inhibition*, *Monitoring* und *Strategieanwendung (inkl. set-shifting)*, sondern auch auf der Integrität lexikalischer und semantischer Verarbeitungsprozesse bezüglich Speicherung, Abruf und Enkodierung (Davis, 2007; Pekkala, 2012). Mit der Evaluation von Wortgenerierungsleistungen sind daher nicht nur Aussagen über Beeinträchtigungen in den Exekutivfunktionen, sondern darüber hinaus auch bezüglich semantischer und lexikalischer Defizite möglich. Wortgenerierungsaufgaben überprüfen Leistungen an der Schnittstelle sprachlicher und kognitiver Fähigkeiten, wodurch sich die hohe Relevanz der Evaluation von Leistungen bei der Wortgenerierung in Bezug auf die Identifizierung kommunikativ-kognitiver Defizite ergibt.

2.5.1 Evaluation von Wortgenerierungsleistungen

Zur Evaluation von Wortgenerierungsleistungen wird traditionell die *Anzahl korrekter Nennungen* in einem vorgegebenen Zeitintervall herangezogen (Aschenbrenner et al., 2000; Paulsen et al., 1996). Die Einschätzung der Leistung erfolgt meist anhand von Prozenträngen. Das heißt, die *Anzahl korrekter Nennungen* wird in Prozentrangwerte transferiert, welche einen Vergleich der individuellen Leistung mit einer Referenzgruppe ermöglichen. Dabei wird bei einem *Prozentrang* < 16 von einer unterdurchschnittlichen und bei einem *Prozentrang* > 84 von einer überdurchschnittlichen Leistung ausgegangen (Aschenbrenner et al., 2000; Fisseni, 2004).²¹ Die *Anzahl korrekter Nennungen* spiegelt als klassisch quantitatives Maß die Korrektheit der Ausführung der Aufgabe wider; eine reduzierte *Anzahl korrekter Nennungen* entspricht dabei einer schlechteren Leistung bei der Wortgenerierung. Wodurch diese reduzierte Wortgenerierungsleistung bedingt ist, geht aus der *Anzahl korrekter Nennungen* jedoch nicht hervor (Drechsler, 2007).

Reduzierte Wortgenerierungsleistungen werden für diverse neurologische und psychiatrische Erkrankungen unterschiedlicher Ätiologien beschrieben, so zum Beispiel für Aphasie (Basso et al., 1997; Kim et al., 2011), Demenz (Araujo et al., 2011), Multiple Sklerose (Beatty, 2002; Friend et al., 1999;

²¹Dieses folgt dem Kriterium, dass durchschnittliche Leistungen im Bereich *Mittelwert* \pm *eine Standardabweichung* liegen (Fisseni, 2004).

Henry & Beatty, 2006), Schizophrenie (Paulsen et al., 1996; Rossell, 2006) oder kommunikativ-kognitive Beeinträchtigungen (Bittner & Crowe, 2006; Martin & McDonald, 2003). Solche quantitativen Leistungsunterschiede finden sich jedoch nicht immer. Insbesondere bei der Dissoziation zwischen verschiedenen Patientengruppen (z.B. bei der Dissoziation zwischen kortikaler und subkortikaler Demenz, vgl. Possin, Vincent Filoteo, Roesch, Zizak, Rilling & Davis, 2005) aber auch bei der Abgrenzung sprachgesunder Leistungen von vergleichsweise subtilen Beeinträchtigungen (Gaspers et al., 2012; Troyer & Moscovitch, 2006) lassen sich nicht immer quantitative Unterschiede finden. Daher kann eine Berücksichtigung qualitativer Maße zur Leistungsbeurteilung erforderlich sein (Possin et al., 2005). Bereits Troyer, Moscovitch und Winocur (1997) wiesen darauf hin, dass es sich bei der Wortgenerierung um eine multifaktorielle Aufgabe handelt. Die Evaluation eines einzigen quantitativen Maßes reicht, nach Meinung der Autoren, für die Charakterisierung einer spezifischen Wortgenerierungsleistung nicht aus.

”(...) fluency is a multifactorial task, and this score [number of correct responses] alone may not fully capture all of the important aspects of any given participant’s performance” (Troyer et al., 1997, S. 138).

Darüber hinaus gibt es Hinweise darauf, dass die *Anzahl korrekter Nennungen* zwar gut geeignet ist, zwischen Patienten mit sprachsystematischen Defiziten und Sprachgesunden zu dissoziieren, jedoch ein weniger geeignetes Maß für die Abgrenzung anderer Störungsbilder oder spezifischer Subgruppen darstellt (Basso et al., 1997; Gaspers et al., 2012; Grossman, 1981). In diesem Zusammenhang gewinnt die Analyse zusätzlicher qualitativer Maße an Bedeutung. Eine Auseinandersetzung mit qualitativen Leistungen in verschiedenen Wortgenerierungsaufgaben kann Informationen liefern, die zu einem grundlegenden Verständnis der Störung beitragen. Solche qualitative Evaluationen von Wortgenerierungsleistungen können beispielsweise sublexikalische, lexikalische, semantische oder auch zeitliche Charakteristika berücksichtigen (Adams et al., 1989; Bittner & Crowe, 2006; Crowe, 1998; Gaspers et al., 2012; Nickels & Howard, 2004).

Häufig befassen sich qualitative Auswertungen mit einer Analyse der Häufigkeit und Art der auftretenden Fehler (Aschenbrenner et al., 2000; Crowe, 1992; Pekkala, 2012), der Clustergröße, das heißt der durchschnittlichen Anzahl von Wörtern innerhalb von lexikalischen oder semantischen Subkategorien (*Clustering*) oder der Anzahl von *Switches* zwischen Subkategorien, das heißt der Fähigkeit zwischen Subkategorien zu alternieren (Beatty et al., 1997; Kavé et al., 2011; Troyer et al., 1997). Hier wird unter anderem kontrovers diskutiert, ob *Clustering* und *Switching* dissoziierbare kognitiv-exekutive Funktionen widerspiegeln, wobei *Clustering* als eher automatische und *Switching* als strategische Komponente gesehen wird (Mayr,

2002; Troyer et al., 1997, 1998; Troyer & Moscovitch, 2006; Unsworth et al., 2011). Spezifische Analysen von Fehlertypen erlauben darüber hinaus Rückschlüsse auf die zugrunde liegende Störung, da das Auftreten bestimmter Fehlermuster mit unterschiedlichen Störungen assoziiert sein kann (Butters et al., 1998; Pekkala, 2012; Possin et al., 2005). Insbesondere verschiedene Formen von Perseverationen oder Regelbrüchen werden mit spezifischen neurogenen Störungsbildern in Verbindung gebracht (Drake et al., 2006; Pekkala, Albert, Spiro III & Erkinjuntti, 2008; Possin et al., 2005). Darüber hinaus wird bei einer eher qualitativ ausgerichteten Evaluation die Anzahl von Nennungen in verschiedenen Zeitintervallen gezählt, das heißt es wird beispielsweise ein Vergleich der Anzahl der Nennungen in 15 oder 30 Sekunden Intervallen vorgenommen (Adams et al., 1989; Bitter & Crowe, 2006; Hall, Harvey, Vo & O'Bryant, 2011; Raboutet, Souz on, Corsini, Rodrigues, Langevin & Kaoua, 2010). Zudem wird von einer Analyse zeitlicher Charakteristika, wie etwa einer Bestimmung der Pausenlange zwischen zwei konsekutiven Nennungen, berichtet (Gaspers et al., 2012; Koren, Kofman, & Berger, 2005; Rosen et al., 2005). Unterschiede in Bezug auf weitere semantische (Adams et al., 1989; Grossman, 1981), lexikalische (Crowe, 1998; Ellis, Miller, & Sin, 1983; Nickels & Howard, 1995) oder sublexikalische (Nickels & Howard, 2004; Pekkala, 2012) Auswertungsparameter werden in der Literatur zwar beschrieben, aber bei der Analyse der Wortgenerierungsfahigkeit bislang nicht hinreichend berucksichtigt.

Zur Erfassung von Wortgenerierungsleistungen existieren verschiedene Verfahren. Im deutschsprachigen Raum steht der Regensburger Wortflussigkeitstest von Aschenbrenner und Kollegen (2000) als standardisiertes und normiertes Testverfahren zur uberprufung von Wortgenerierungsleistungen zur Verfugung. Neben funf simplen lexikalischen (Wortern mit den Anfangsbuchstaben *S*, *P*, *M*, *K* oder *B*) und funf simplen semantischen Wortgenerierungsaufgaben (Wortern aus den Kategorien *Vornamen*, *Tiere*, *Lebensmittel*, *Berufe* oder *Hobbys*) umfasst der Regensburger Wortflussigkeitstest auch je zwei intradimensionale Wechsel-Aufgaben (lexikalisch: *Wechsel G/R* oder *H/T*, semantisch: *Wechsel Sportarten/Fruchte* oder *Kleidungsstucke/Blumen*) zur Erfassung komplexerer Wortgenerierungsleistungen. Der Test liefert unter anderem Normtabellen fur Erwachsene in funf verschiedenen Altersgruppen, jeweils fur eine Testdauer von 60 und 120 Sekunden. Daruber hinaus liegen geschlechts- und bildungsspezifische Tabellen zur Leistungsbeurteilung vor.

Im englischsprachigen Raum wird fur die uberprufung lexikalischer Wortgenerierungsleistungen der *Controlled Oral Word Association-Test* (COWA/COWAT) herangezogen (Barry et al., 2008). In seiner ursprunglichen Form beinhaltet dieser das Produzieren von Wortern mit den Anfangsbuchstaben *C*, *F* und *L* (Version 1) oder *P*, *R* und *W* (Version 2). In neuerer Zeit werden vor allem die Anfangsbuchstaben *F*, *A* und *S* zur Elizitierung lexikalischer

Wortgenerierungsleistungen verwendet (Barry, Bates, & Labouvie, 2008). Zur Erhebung semantischer Wortgenerierungsfähigkeit werden im englischsprachigen Raum verschiedene Objektkategorien herangezogen. Am häufigsten werden die sogenannte *animal naming fluency*, das heißt das Produzieren von Wörtern aus der Kategorie *Tiere*, oder die assoziativ etwas weiter gefasste *Supermarktaufgabe*, bei der möglichst viele Dinge genannt werden sollen, die man in einem Supermarkt kaufen kann, evaluiert (vgl. Pekkala, 2012; Tombaugh, Kozak, & Rees, 1999; Troyer & Moscovitch, 2006; Unsworth, Spillers, & Brewer, 2011).

Etablierte, standardisierte Verfahren zur Erhebung von Leistungen bei der Verbgenerierung gibt es meines Wissens bislang nicht.

In der vorliegenden Arbeit sollen quantitative und qualitative Aspekte der Evaluation von Wortgenerierungsleistungen unter Berücksichtigung verschiedener Methoden zur Diagnose von kommunikativ-kognitiven Beeinträchtigungen herangezogen werden. Dabei stellt die Anwendung von *machine learning*-Techniken eine Herangehensweise dar. Aufgrund ihres innovativen Charakters sollen im Folgenden die theoretischen Grundlagen des *machine learning* eingeführt werden.

2.6 *machine learning*

Eine neuere Methodologie zur Identifizierung spezifischer Störungsbilder liefert der Einsatz von *machine learning*-Techniken. Der Begriff *machine learning* sowie die dahinter stehende Methode stammen aus dem Bereich der Informatik. Mitchell (1997) definiert den Begriff *machine learning* wie folgt:

”A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.” (Mitchell, 1997, S. 2)

Vereinfacht gesprochen heißt das: Ein künstliches System kann durch Erfahrungen, die es aus Beispielen (Trainingsdaten) gewinnt, Wissen aufbauen. Es lernt Gesetzmäßigkeiten und Zusammenhänge zu erkennen und kann nach Abschluss einer Trainingsphase daraus verallgemeinern (Bishop, 2006; Nilsson, 1998). *machine learning* kann damit auch für verschiedene Klassifikationsaufgaben verwendet werden. So zum Beispiel bei der Dissoziation zwischen verschiedenen Patientengruppen anhand von Daten, die Leistungen in verschiedenen kognitiven oder linguistischen Tests abbilden (Järvelin & Juhola, 2011; Quintana et al., 2012). *machine learning*-Techniken können eine sinnvolle Methode bei der Diagnosestellung sein, werden aber auch in

vielen anderen Bereichen wie etwa bei SPAM-Filtern oder bei der Spracherkennung eingesetzt.

Es werden verschiedene Lernalgorithmen unterschieden. Im Wesentlichen sind das *supervised learning*, *unsupervised learning* und *reinforcement learning* (Bishop, 2006; Mitchell, 1997; Nilsson, 1998). Beim *supervised learning* wird auf der Basis bestehender bzw. vorab definierter Kategorien (z.B. Sprachgesund vs. Aphasie) trainiert. Der Algorithmus lernt anhand von Beispielen spezifische Charakteristika einer bestimmten Kategorie und kann den bestehenden Kategorien im Anschluss an den Lernprozess neue Daten zuordnen. Beim *unsupervised learning* besteht das Trainingsset demgegenüber aus Inputdaten für die keine vorab definierten Kategorien existieren. Es ermöglicht in einer großen ungruppierten Datenmenge das Erkennen natürlicher Muster, anhand derer die Daten sinnvoll kategorisiert werden können (auch als *data mining* bezeichnet; vgl. Bishop 2006, Nilsson, 1998). Beim *reinforcement learning* lernt ein System auf Basis von Feedback (Belohnung) welches es aus seiner Umgebung erhält, in einer spezifischen Situation angemessen zu reagieren (Bishop, 2006; Nilsson, 1998).

Klassifikation

Für die vorliegende Arbeit ist insbesondere der Klassifikationsaspekt von Bedeutung. Klassifikationen werden im *machine learning*, wie bereits angedeutet, dem *supervised learning* zugeordnet. Lernen geschieht hierbei auf Basis eines Trainingsdatensets, welches korrekt identifizierte Zuordnungen bzw. Kategorien (z.B. Aphasie vs. Sprachgesund) enthält. Der Algorithmus lernt anhand festgelegter Kategorien durch einen Vergleich von Input-Output-Funktionen. Dabei sind Input- und Output-Daten durch Vektoren repräsentiert. Das Trainingsdatenset besteht also aus Input-Daten, deren Kategoriezugehörigkeit (*output vector*) bekannt ist. Die Trainingsdaten werden genutzt, um die Parameter eines adaptiven Modells zu verbessern. Das Ziel ist dabei, jedem Input eine spezifische Kategorie zuzuordnen (Bishop, 2006; Marsland, 2009).

Die Kategorien der Beispiele (auch *Instanzen* genannt) im Trainingsdatenset sind bekannt. Die zu klassifizierenden Instanzen werden durch verschiedene Messdaten beschrieben, das heißt, die Input-Daten sind durch spezifische Merkmale (Variablen) charakterisiert. Diese bilden die Grundlage für das Erkennen von Zusammenhängen und Mustern. Ergebnis des *machine learning*-Prozesses ist, dass jedem neuen Input (d.h. jeder neuen Instanz) ein spezifischer Output-Vektor (d.h. eine Kategorie) zugewiesen wird. Auf welcher Basis diese Zuordnung erfolgt, entscheidet sich in der Trainingsphase. Anhand der Trainingsdaten und ihrer spezifischen Charakteristika lernt das System die Kategoriezuordnungen. Ist das Modell nach Ende die-

ser Lernphase trainiert, kann es auch neuem Input Kategorien zuweisen. Diese neuen Daten sind Teil eines Testsets. Ist das Modell in der Lage neue Beispiele korrekt zu klassifizieren, spricht man von Generalisierung (Bishop, 2006; Marsland, 2009).

Vor der Klassifikation werden die Input-Daten in der Regel vorverarbeitet um die Variabilität zu verringern und die Daten vergleichbar zu machen. Dies soll die Mustererkennung vereinfachen (Jain, Duin, & Mao, 2000). Während der Klassifikation wird für die Input-Vektoren entschieden, zu welcher der bestehenden Kategorien sie sich zuordnen lassen. Die Entscheidung wird auf Basis der Beispiele für jede Kategorie aus dem jeweiligen Trainingsdatenset getroffen.

Eine Klassifikation kann mit verschiedenen Klassifikatoren bzw. auf Basis unterschiedlicher Klassifikationsalgorithmen erfolgen, welche jeweils auf spezifischen mathematischen Operationen beruhen (Bishop, 2006; Marsland, 2009). Allen Klassifikationen ist gemein, dass mit Ihnen Entscheidungsgrenzen gefunden werden sollen, die es ermöglichen, verschiedene Gruppen oder Kategorien anhand spezifizierender Variablen voneinander zu separieren (Marsland, 2009). Die verschiedenen Klassifikatoren unterscheiden sich bezüglich ihrer Algorithmen und deren Komplexität. Die Wahl eines Klassifikators ist unter anderem abhängig von den verwendeten Daten sowie von dem zu erwartenden Rechenaufwand (Järvelin & Juhola, 2011; Marsland, 2009). Zu den am meisten beschriebenen und für die Klassifikation verwendeten Klassifikatoren zählen: (i) *Support Vector Machines* (SVM) als *maximum margin* Klassifikator; (ii) *Artificial Neural Networks* (d.h. künstliche neuronale Netze) wie das *Multi Layer Perceptron* (MLP); (iii) Entscheidungsbäume, die auf Basis einer wiederholten Selektion einzelner Variablen operieren; (iv) *k-nearest neighbor* (KNN), der als relativ simpler Klassifikator auf dem Prinzip der Ähnlichkeit beruht, oder (v) statistische Modellierung, wie z.B. der auf Grundlage des Bayes Theorem entwickelte probabilistische Klassifikator *Naive Bayes* etc. (vgl. Bishop, 2006; Marsland, 2009; Mitchell, 1997). Eine detaillierte Übersicht verschiedener Klassifikationsalgorithmen findet sich in einem Review von Jain und Kollegen (2000).

Zur weiteren Verbesserung des Klassifikationsergebnisses kann es darüber hinaus sinnvoll sein, verschiedene Klassifikatoren miteinander zu kombinieren (Jain et al., 2000). Dabei werden die Ergebnisse individueller Klassifikatoren anhand eines Meta-Klassifikators miteinander kombiniert, der dann eine Gesamtentscheidung trifft (Jain et al., 2000). Die Art und Weise, in der die Ergebnisse der einzelnen Klassifikatoren miteinander kombiniert werden, kann dabei variieren (Jain et al., 2000). Unterschieden werden hier im Wesentlichen parallele, kaskadische/serielle oder hierarchische Möglichkeiten der Kombination durch einen Meta-Klassifikator (Jain et al., 2000).

Die parallele Verarbeitung stellt dabei die am häufigsten verwendete Methode der Meta-Klassifikation dar. Hierbei werden die individuellen Klassifikatoren zuerst separat angewendet und ihre Ergebnisse erst im Anschluss kombiniert. Demgegenüber werden bei der kaskadischen Meta-Klassifikation die Ergebnisse der individuellen Klassifikatoren einer linearen Sequenz folgend kombiniert; eine hierarchische Meta-Klassifikation ist vom Prinzip einem Entscheidungsbaum ähnlich (Jain et al., 2000).

Ein weiterer wichtiger Aspekt ist die Validierung der Klassifikationsmethode. Diese ermöglicht eine Einschätzung darüber, wie gut das trainierte Modell funktioniert (Marsland, 2009; Nilsson, 1998). Zur Validierung des Modells wird häufig eine Kreuzvalidierung herangezogen. Steht nur ein sehr geringer Datensatz zur Verfügung, bei dem das Zurückhalten eines Validierungsdatensets (zusätzlich zum Trainings- und Testset) die Gefahr birgt, auf den verbliebenen Daten nicht ausreichend gut trainieren zu können, kann auf eine mehrfache Kreuzvalidierung (*multi-fold cross validation*, wie z.B. die *10-fold cross validation* bei Järvelin & und Juhola, 2011) zurückgegriffen werden. Dazu wird das Datenset in verschiedene kleinere Sets aufgeteilt. Ein Set wird dann als Validierungsdatenset zurückbehalten, auf den verbliebenen Sets wird trainiert. Dieser Prozess wird mit allen kleineren Sets durchgeführt bis jedes einmal als Validierungsdatenset genutzt wurde (Marsland, 2009).

Die Leistung eines Klassifikators bzw. die Güte des Klassifikationsergebnisses wird anhand verschiedener Maße bewertet. Das Wichtigste ist dabei sicherlich die *accuracy*, das heißt die Treffergenauigkeit oder Klassifikationsgenauigkeit. Sie gibt den Anteil korrekt klassifizierter Instanzen wieder. Darüber hinaus existieren weitere Evaluationsmaße, ähnlich denen, die auch bei der Einschätzung klinischer Testverfahren verwendet werden (Bortz & Lienert, 2008). So sind Aussagen über Spezifität und Sensitivität oder falsch positive/negative Klassifikationen möglich (Bortz & Lienert, 2008).

2.6.1 machine learning in der klinischen Anwendung

Verschiedene Studien konnten zeigen, dass *machine learning*-Techniken zur Evaluation klinischer Testverfahren und diagnostischer Merkmale in verschiedenen klinischen Populationen sinnvoll eingesetzt werden können (Costafreda et al., 2011; Fraser et al., 2012; Gaspers et al., 2012; Järvelin, 2008; Järvelin & Juhola, 2011; Quintana et al., 2012; Wilson et al., 2009). Einige der Studien mit Relevanz für die vorliegende Dissertation sollen an dieser Stelle kurz aufgeführt werden.

Järvelin und Juhola (2011) untersuchten die Anwendung von acht Klas-

sifikatoren bei verschiedenen Klassifikationsproblemen aus dem Bereich der Aphasie. Die Autoren analysierten zum einen Daten aus Tests zum konfrontativen Benennen (eliziert mit verschiedenen Verfahren) und untersuchten wie gut verschiedene Klassifikatoren zwischen aphasischen und nicht-aphasischen Patienten sowie zwischen Patienten mit Alzheimer und vaskulärer Demenz dissoziieren können. Zum anderen verwendeten sie Daten aus dem Aachener Aphasie Test (Huber et al., 1983), um zu evaluieren wie gut die Klassifikatoren zwischen den verschiedenen Aphasiesyndromen dissoziieren können. Die Analysen ergaben, dass keiner der Klassifikatoren bei allen Klassifikationsaufgaben gute Ergebnisse erzielte. Das zeigt, dass die Wahl eines Klassifikators immer für ein spezifisches Datenset getroffen werden sollte und es keinen universellen Klassifikator für Klassifikationsprobleme aus dem Bereich der Aphasie zu geben scheint.

Costafreda und Kollegen (2011) verwendeten *machine learning*-Techniken, um Patienten mit Schizophrenie, bipolarer Störung und Sprachgesunde zu klassifizieren. Sie analysierten Muster neuronaler Aktivierung bei der Wortgenerierung mit dem Ziel, diese als Marker für die Diagnosestellung nutzen zu können. Die Analysen ergaben distinkte Aktivierungsmuster, welche relevante Informationen für die Diagnosestellung lieferten. Über 90% der Patienten mit Schizophrenie konnten anhand des Musters neuronaler Aktivierung bei der Wortgenerierung korrekt klassifiziert werden.

Die Analysen von Fraser et al. (2012) befassten sich mit der Quantifizierung von subtilen sprachlichen Beeinträchtigungen bei Patienten mit neurodegenerativen Erkrankungen. Dabei evaluierten und klassifizierten sie Patienten mit primär progressiver Aphasie, semantischer Demenz und Sprachgesunde anhand syntaktischer und semantischer Variablen beim Nacherzählen eines Märchens. Bei der Analyse relevanter Variablen konnten die Autoren Hinweise darauf finden, dass insbesondere die Wortfrequenz ein geeignetes Parameter für die Dissoziation zwischen den drei Gruppen darstellt.

Quintana und Kollegen (2012) nutzten neuronale Netze, um zwischen Patienten mit leichten kognitiven Beeinträchtigungen und Patienten mit Alzheimer Demenz zu dissoziieren. Als sensitivste Variablen für die Klassifikation konnten semantische Wortgenerierungsleistungen, Arbeitsgedächtniskapazität und Leistungen des episodischen Gedächtnisses identifiziert werden. Der Einsatz neuronaler Netze erlaubte eine reliable Dissoziation zwischen den Patientengruppen und neurologisch gesunden Personen und wurde daher von den Autoren als geeignete Methode zur Analyse neuropsychologischer Beeinträchtigungen vorgeschlagen.

Wilson und Kollegen (2009) untersuchten verschiedene Varianten primär progressiver Aphasie mit *machine learning*-Techniken. Die Autoren evaluierten hierzu sowohl linguistische Variablen als auch Bildgebungsbefunde bei Patienten mit progressiver nicht-flüssiger Aphasie, semantischer Demenz und logopenisch progressiver Aphasie sowie Sprachgesunden. Bei Verwendung eines SVM Klassifikators und Kreuzvalidierung konnten über 90% der

Patienten korrekt klassifiziert werden. Die besten Klassifikationsergebnisse wurden erzielt, wenn Modelle verwendet wurden, die eine Kombination aus linguistischen und Bildgebungsdaten verwendeten. Die Autoren sehen in der von ihnen verwendeten Analyse das Potenzial, sie als Methode für die Differentialdiagnose bei primär progressiver Aphasie einsetzen zu können, mit dem Ziel die zugrunde liegende Ätiologien aufzudecken.

Die aufgeführten Studien zeigen, dass der Einsatz von *machine learning*-Techniken eine sinnvolle Erweiterung traditioneller Analysemethoden für die Diagnose verschiedener Beeinträchtigungen in Sprache, Kognition oder Kommunikation darstellen kann.

2.7 Fragestellungen

Im Folgenden sollen nun drei empirische Studien vorgestellt werden, die das Herzstück dieser Dissertation darstellen. Die erste Studie, im nun folgenden Kapitel, beschreibt die Untersuchung quantitativer und qualitativer Charakteristika der Wortgenerierungsleistung bei Patienten mit Multipler Sklerose. Daran anschließend folgt in Kapitel 4 mit der zweiten Studie eine Evaluation von Wortgenerierungsleistungen zur Klassifikation von Sprach- und Kommunikationsstörungen mit *machine learning*-Techniken. Die dritte Studie in Kapitel 5 widmet sich der Exploration einer Methodologie zur Evaluation von Wortgenerierungsleistungen bei Patienten mit kommunikativ-kognitiven Defiziten. Studienübergreifend sind dabei die folgenden Fragen von Bedeutung:

- (a) Ist die Evaluation von Wortgenerierungsleistungen zur Diagnose kommunikativ-kognitiver Defizite und anderer Sprach- und Kommunikationsstörungen geeignet?
- (b) Welche Auswertungsparameter ermöglichen eine reliable Diagnosestellung? Unterscheiden sich die Auswertungsparameter für die Diagnose verschiedener Störungsbilder?
- (c) Welche Rückschlüsse können aus reduzierten Wortgenerierungsleistungen oder spezifischen Leistungsunterschieden gezogen werden?
- (d) Welche der verwendeten Methodologien kann für die Evaluation von Wortgenerierungsleistungen sinnvoll eingesetzt werden?

Diese Fragen sollen in einer abschließenden Diskussion in Kapitel 6 wieder aufgegriffen und auf Basis der Befunde aus den drei Studien diskutiert werden.

Kapitel 3

Evaluation quantitativer und qualitativer Charakteristika der Wortgenerierungsleistung bei Multipler Sklerose

3.1 Einleitung & Fragestellungen

Als eines der sensitivsten Maße zur Evaluation kognitiver Leistungen bei Patienten mit Multipler Sklerose wird die Wortgenerierungsfähigkeit beschrieben (Henry & Beatty, 2006). Noch bestehen jedoch Kontroversen bezüglich des zugrunde liegenden Leistungsdefizits bei Patienten mit Multipler Sklerose. Ergebnisse bisheriger Studien weisen in keine eindeutige Richtung (Beatty, 2002; Henry & Beatty, 2006; Zakzanis, 2000). Das heißt, es besteht weiterhin die Frage, ob primär Defizite in den Exekutivfunktionen oder Beeinträchtigungen der Integrität semantischer Netzwerke ursächlich für die Wortgenerierungsdefizite bei Patienten mit Multipler Sklerose sind (Henry & Beatty, 2006). Von stärker beeinträchtigten Leistungen bei der semantischen Wortgenerierung (Sepulcre et al., 2011; Zakzanis, 2000) wird ebenso berichtet wie von gleichermaßen beeinträchtigten Leistungen bei lexikalischer und semantischer Wortgenerierung (Beatty et al., 1989; Beatty, 2002; Kraus et al., 2005; Messinis et al., 2013), sowie von einer Dominanz lexikalischer Wortgenerierungsdefizite (Henry & Beatty, 2006; Nocentini et al., 2001).

Aufgrund der Annahme, dass der Wortgenerierungsfähigkeit in lexikalischen und semantischen Aufgaben zum Teil unterschiedliche Verarbeitungsprozesse zugrunde liegen (Baldo et al., 2010), können Defizite, die sich durch

eine signifikant geringere *Anzahl korrekter Nennungen* in einzelnen oder verschiedenen Aufgaben zeigen, auf spezifische Störungsschwerpunkte hinweisen. Eine verminderte Wortgenerierungsleistung in semantischen Aufgaben bei erhaltener lexikalischer Wortgenerierungskompetenz kann beispielsweise als Hinweis auf ein semantisches Defizit gewertet werden, während ähnliche Leistungen in semantischen und lexikalischen bzw. schlechtere Leistungen in lexikalischen Wortgenerierungsaufgaben ein Defizit in den Exekutivfunktionen nahelegen (vgl. Abschnitt 2.3; Henry & Beatty, 2006; Henry & Phillips 2006). Die Evaluation von Leistungen in kognitiv-exekutiv anspruchsvolleren Wechsel-Aufgaben sowie bei der Verbgenerierung kann darüber hinaus zur Charakterisierung des Leistungsdefizits beitragen. Untersuchungen, die sich eingehend mit der Analyse von Wortgenerierungsleistungen in Wechsel-Aufgaben und der Verbgenerierung bei Multipler Sklerose befassen, gibt es bislang nicht.

Darüber hinaus wird kontrovers diskutiert, ob sich die verschiedenen Verlaufsformen einer Multiplen Sklerose anhand kognitiver Leistungsprofile voneinander abgrenzen lassen (vgl. Abschnitt 2.4.3; Drake et al. 2006; Kraus et al. 2005; Nocentini et al. 2001; Ruet et al. 2013 oder Wishart & Sharpe 1997). Während einige Autoren quantitative Unterschiede in semantischen und lexikalischen Wortgenerierungsaufgaben zwischen den Verlaufsformen einer Multiplen Sklerose beschreiben (Beatty et al., 1989; Henry & Beatty, 2006), finden andere lediglich Unterschiede in der Leistungsqualität (Messinis et al., 2013). Eine zuverlässige Identifizierung und aussagekräftige Beschreibung der Defizite sowie des zugrunde liegenden Leistungsdefizits sind Voraussetzung für eine fundierte Therapieplanung (Tesak, 2006; Wehmeyer & Grötzbach, 2006). Daher soll im Rahmen dieser Studie die Wortgenerierungsfähigkeit bei Patienten mit schubförmig-remittierender und Patienten mit chronisch-progredienter Multipler Sklerose sowohl quantitativ als auch qualitativ in verschiedenen Wortgenerierungsaufgaben im Vergleich zur Leistung Sprachgesunder evaluiert werden. Das Ziel ist dabei, einer Beschreibung des zugrunde liegenden Defizits näher zu kommen.

Traditionell wird bei der klinischen Evaluation von Wortgenerierungsleistungen die *Anzahl korrekter Nennungen* ausgezählt und als quantitatives Leistungsmaß verwendet (Aschenbrenner et al., 2000; Paulsen et al., 1996; Troyer et al., 1997). Quantitative Leistungsunterschiede finden sich jedoch nicht immer und ein Einbeziehen qualitativer Aspekte kann zu einem besseren Verständnis des Leistungsdefizits beitragen (Possin et al., 2005). Eine Möglichkeit der qualitativen Beurteilung einer Leistung bietet die Evaluation der auftretenden Fehler. Treten Fehler gehäuft auf, das heißt in einem stärkerem Ausmaß als bei Sprachgesunden, können Rückschlüsse auf spezifische Leistungsdefizite möglich sein (Aschenbrenner et al., 2000; Crowe, 1992). Auch die verschiedenen Fehlertypen legen bestimmte Interpretatio-

nen im Hinblick auf die zugrunde liegende Störung nahe und können darüber hinaus zur Abgrenzung verschiedener Störungsbilder beitragen (Butters et al., 1998; Possin et al., 2005).

Die existierende Terminologie möglicher Fehlertypen ist nicht ganz einheitlich bzw. es fehlen häufig klare Abgrenzungen einzelner Fehlertypen (Aschenbrenner et al., 2000; Possin et al., 2005). Die am meisten verwendete Unterscheidung zur Beschreibung auftretender Fehlertypen besteht in der Abgrenzung von *Perseverationen* (auch: *Repetitionen* genannt, vgl. Aschenbrenner et al., 2000) und *Regelbrüchen* (Aschenbrenner et al., 2000; Pekkala, 2012; Possin et al., 2005). Eine weit verbreitete Taxonomie ist dabei die von Sandson und Albert (1984). Die Autoren beschreiben drei Formen von Perseverationen, welche sich sowohl hinsichtlich der Phänomenologie, als auch bezüglich der zugrunde liegenden Prozesse sowie der neuroanatomischen Lokalisationen abgrenzen lassen: *continuous perseveration*, *recurrent perseveration* und *stuck in set perseveration* (Sandson & Albert, 1984).

Continuous perseverations werden häufig im Zusammenhang mit Läsionen der Basalganglien beschrieben und bezeichnen eine unangemessene zeitliche Ausdehnung und Wiederholung ohne Unterbrechung, welche unter anderem auf ein Defizit in der motorischen Ausführung zurückgeführt wird (Possin et al., 2005; Sandson & Albert, 1984). Bei der Evaluation der vorliegenden Wortgenerierungsdaten spielen *continuous perseverations* eine untergeordnete Rolle und werden daher nicht weiter berücksichtigt.

Als *recurrent perseveration* wird die unbewusste Wiederholung einer vorherigen Antwort nach einem Stimulus bezeichnet (Possin et al., 2005; Sandson & Albert, 1984). Diese Form der Perseveration kann sowohl direkt auf den Stimulus folgend als auch zeitverzögert nach einigen weiteren Stimuli auftreten. *Recurrent Perseverations* werden häufig bei Sprachstörungen nach linkshemispärischen Läsionen, das heißt im Rahmen von Aphasien, aber auch bei der Alzheimer Demenz berichtet und äußern sich beispielsweise beim Bildbenennen (Sandson & Albert, 1984). Ihr Auftreten wird unter anderem im Zusammenhang mit beeinträchtigten Gedächtnisspeicherprozessen bzw. als Defizit bei der Inhibition zuvor aktivierter Wörter diskutiert (Aschenbrenner et al., 2000; Moses, Sheard & Nickels, 2007; Possin et al., 2005). *Recurrent perseverations* sind definitionsgemäß mit den im Regensburger Wortflüssigkeitstest als Repetition bezeichneten Fehlern gleichzusetzen (Aschenbrenner et al., 2000).

Eine *stuck in set perseveration* wird als Haftenbleiben an einer bestimmten Aufgabe bzw. einem Konzept beschrieben, das heißt, ein Wechsel zwischen verschiedenen Konzepten gelingt nicht. Diese Form der Perseveration wird insbesondere mit einer Störung der Exekutivfunktionen nach unilateralen links- oder rechtshemisphärischen Frontalhirnläsionen in Verbindung gebracht und kann Ausdruck einer Beeinträchtigung der kognitiven Flexibilität, sowie von Defiziten beim Wechsel des Aufmerksamkeitsfokus

sein (Sandson & Albert, 1984). *Stuck in set perseverations* werden modelltheoretisch auch in Verbindung mit einer Störung im *Supervisory Attentional System* (Aschenbrenner et al., 2000; Norman & Shallice, 1980; Shallice et al., 1996) beschrieben. Durch ein defizitäres *Supervisory Attentional System* kann nicht von routinierten Handlungen bzw. Schemata abgewichen werden; eine flexible Anpassung sowie Monitoring sind nicht möglich. Durch das Haftenbleiben an für die erfolgreiche Lösung der Aufgabe unangemessenen Routinen kommt es zum Auftreten der *stuck in set perseverations*. Die im Regensburger Wortflüssigkeitstest beschriebenen Kategorieperseverationen bei den Wechsel-Aufgaben können als Form einer *stuck in set perseveration* gesehen werden (Aschenbrenner et al., 2000).

Als Regelbrüche (auch *intrusion* genannt) sind, bezogen auf die Evaluation von Wortgenerierungsdaten, solche Nennungen zu bezeichnen, die entgegen der Aufgabenstellung produziert werden. Das heißt, alle Nennungen, die nicht den vorgegebenen Regeln entsprechen und/oder in keinem Zusammenhang mit der Aufgabenstellung stehen (z.B. das Nennen von Eigennamen und Pseudowörtern oder unrelatierter Wörter). Regelbrüche sind bei Sprachgesunden selten zu beobachten und werden bei neurologisch bedingten kommunikativen und kognitiven Störungen als Hinweis auf Defizite im Bereich der Exekutivfunktionen gewertet (Pekkala, 2012). Regelbrüche indizieren ein Defizit bei der Inhibition unpassender Nennungen und werden darüber hinaus als Beeinträchtigung des Zugriffs auf, sowie beim Abruf und der Integration von lexikalischem und semantischem Wissen diskutiert. Bei semantischen Wortgenerierungsaufgaben gelten Regelbrüche darüber hinaus als Indiz für unscharf definierte semantische Kategorien (Possin et al., 2005). Das Auftreten von Regelbrüchen und *stuck in set perseverations* wird als Indikator subkortikaler Läsionen beschrieben (Butters et al., 1998).

Studien, die sich mit dem Auftreten von Regelbrüchen und Perseverationen bei Patienten mit Multipler Sklerose befassen, fokussieren vor allem Aufgaben zur Überprüfung spezifischer Gedächtnisleistungen oder allgemeiner kognitiver Funktionen (Butters et al., 1998; Drake et al., 2006). Butters und Kollegen (1998) untersuchten kognitive Funktionen bei Patienten mit Multipler Sklerose im Vergleich zu Patienten mit Chorea Huntington und Patienten mit kortikaler Demenz. Obwohl Patienten mit Multipler Sklerose im Vergleich zu den beiden anderen Patientengruppen relativ geringe kognitive Beeinträchtigungen zeigten, konnte eine große Anzahl von *recurrent perseverations* und *stuck in set perseverations* sowie auch von stimulusbezogenen Fehlern (ähnlich semantischen/phonematischen Paraphasien) gefunden werden. Bei den drei untersuchten Gruppen waren insgesamt qualitativ unterschiedliche Fehlerschwerpunkte zu beobachten. Die bei der Evaluation kognitiver Leistungen gefundenen Fehler bei Patienten mit Multipler Sklerose wiesen auf kognitiv-rigide Funktionen hin, wie sie sonst typischerweise bei Patienten mit Morbus Parkinson beschrieben werden (Butters et al., 1998).

Zur Identifizierung demenzieller Beeinträchtigungen bei Patienten mit Morbus Parkinson konnte insbesondere die Evaluation von Leistungen bei der Verbgenerierung beitragen (Piatt et al., 1999a). Der Einsatz von Verbgenerierungsaufgaben könnte daher auch bei Patienten mit Multipler Sklerose einen wertvollen Beitrag zur weiteren Charakterisierung des Leistungsdefizits liefern. Studien, die sich mit der qualitativen Analyse von Perseverationen und Regelbrüchen bei der Überprüfung von Wortgenerierungsleistungen in verschiedenen Aufgaben bei Patienten mit schubförmig-remittierender und chronisch-progredienter Multipler Sklerose befassen, gibt es bisher nicht.

Im Rahmen dieser Studie sollen Wortgenerierungsleistungen von Patienten mit schubförmig-remittierender (MS_SR) und chronisch-progredienter (MS_CP) Multipler Sklerose im Vergleich zur Leistung Sprachgesunder in vier klassischen Wortgenerierungsaufgaben sowie einer Aufgabe zur Verbgenerierung evaluiert werden. Das Ziel ist dabei, der Charakterisierung eines möglichen sprachlichen und/oder kognitiven Leistungsdefizits bei Multipler Sklerose näher zu kommen. Neben einer Analyse der *Anzahl korrekter Nennungen* als Leistungsmaß in den einzelnen Aufgaben, soll auch der Anteil korrekter Nennungen in Relation zur Gesamtzahl der Nennungen (als indirektes Fehlermaß) sowie das Auftreten von Perseverationen und Regelbrüchen berücksichtigt werden. Von Interesse ist insgesamt, ob sich für die drei untersuchten Gruppen spezifische Wortgenerierungsprofile erkennen lassen. Relevante Fragen sind dabei: (i) Unterscheidet sich die Wortgenerierungsleistung der Patienten mit MS_SR und MS_CP von den Leistungen sprachgesunder Probanden generell bzw. spezifisch in den einzelnen Wortgenerierungsaufgaben? (ii) Unterscheidet sich der Anteil korrekter Nennungen (als indirektes Fehlermaß) in den einzelnen Wortgenerierungsaufgaben zwischen Patienten mit MS_SR und MS_CP und den Sprachgesunden? (iii) Unterscheiden sich die drei Gruppen bezüglich des Anteils von Perseverationen und Regelbrüchen in den verschiedenen Wortgenerierungsaufgaben?

3.2 Methode

Die hier beschriebenen Daten sind zum Teil im Rahmen einer größeren Studie erhoben worden, bei der unter anderem auch spontansprachliche Äußerungen sowie Assoziations- und Wortgenerierungsleistungen in weiteren Kategorien erhoben wurden. Erste Ergebnisse der Evaluation einiger Spontanspracheparameter finden sich in Böwering (2012). Ein weiterer Bericht, der sich mit der Darstellung der Relation von Spontansprache- und Wortgenerierungscharakteristika befasst, ist in Vorbereitung (Thiele, Jaecks, Böwering, & Stenneken, in Vorbereitung). Im Folgenden soll nun die Evaluation der Wortgenerierungsleistungen im Vordergrund stehen.

3.2.1 Versuchsteilnehmer

Es werden Daten von insgesamt 32 Versuchsteilnehmern aus drei Gruppen präsentiert. Alle Versuchsteilnehmer waren monolinguale, deutsche Muttersprachler mit normalem oder korrigiertem Seh- und Hörvermögen. Die Datenerhebung erfolgte in Übereinstimmung mit den 1964 in der Helsinki Deklaration festgelegten ethischen Standards. Alle Teilnehmer gaben nach erfolgter Aufklärung ihre Einwilligung zur Teilnahme an der Studie.

Patienten mit MS_SR. Die erste Gruppe bestand aus 10 Patienten (3 Männer, 7 Frauen) mit schubförmig-remittierender Multipler Sklerose, im Alter von 24 bis 51 Jahren ($M = 41.20$, $SD = 8.00$). Die durchschnittliche Krankheitsdauer²² reichte von 17 bis 182 Monate ($M = 84.00$, $SD = 64.91$). Der Bildungsgrad betrug im Mittel 11.50 Schuljahre ($SD = 1.43$, $Min = 10$, $Max = 14$).

Patienten mit MS_CP. Die zweite Gruppe bestand aus 6 Patienten (5 Männer, 1 Frau) mit chronisch-progredienter Multipler Sklerose, im Alter von 44 bis 66 Jahren ($M = 51.50$, $SD = 7.94$). Die durchschnittliche Krankheitsdauer reichte von 69 bis 254 Monate ($M = 133.83$, $SD = 66.68$). Der Bildungsgrad betrug im Mittel 11.50 Schuljahre ($SD = 1.38$, $Min = 10$, $Max = 13$).

Ein nicht-parametrischer Gruppenvergleich für unabhängige Stichproben (Mann-Whitney U Test) ergab, dass sich die mittlere Krankheitsdauer zwischen den Patienten mit MS_SR und MS_CP nicht signifikant voneinander unterschied ($U = 18.00$, $Z = -1.30$, $p = .19$).

Sprachgesunde. Die dritte Gruppe bestand aus 16 neurologisch gesunden Versuchsteilnehmern (5 Männer, 11 Frauen) im Alter von 26 bis 60 Jahren ($M = 44.94$, $SD = 9.07$) und mit einem mittleren Bildungsgrad von 12.19 Schuljahren ($SD = 1.38$, $Min = 10$, $Max = 14$).

Ein nicht-parametrischer Gruppenvergleich für k -unabhängige Stichproben nach Kruskal-Wallis ergab keine signifikanten Unterschiede bezüglich Alter ($\chi^2(2) = 4.91$, $p = .086$) und Bildungsgrad ($\chi^2(2) = 2.52$, $p = .28$) zwischen den drei Gruppen. Auch die Überprüfung hinsichtlich der Geschlechterverteilung in den drei Gruppen mit dem Chi-Quadrat Test nach Pearson erbrachte keinen signifikanten Unterschied zwischen den drei Gruppen ($\chi^2(2) = 5.59$, $p = .06$.)

Zur weiteren Beschreibung der drei Gruppen liegen für den überwiegenden

²²Das heißt: Zeit seit Diagnosestellung zum Zeitpunkt der Testung.

den Teil der Versuchsteilnehmer²³ (MS_SR: $n = 9$, MS_CP: $n = 6$ und SG: $n = 11$) darüber hinaus Ergebnisse aus dem DemTect (Kalbe, Kessler, Calabrese, Smith, Passmore, Brand, & Bullock, 2004) vor, einem Screening für leichte kognitive Beeinträchtigungen bzw. frühe Symptome einer demenziellen Erkrankung. Im DemTect entspricht ein Testwert von 13 bis 18 einer altersgemäßen Leistung; ein Testwert im Bereich von 9 bis 12 weist auf leichte kognitive Beeinträchtigungen hin. Ist der Testwert kleiner/gleich 8 besteht ein Verdacht auf Demenz. Hinweise auf das Vorliegen leichter kognitiver Defizite zeigten sich bei vier Versuchsteilnehmern aus der MS_CP Gruppe (siehe Tabelle 3.1). Alle anderen Versuchsteilnehmer zeigten altersentsprechende Leistungen. Ein nicht-parametrischer Gruppenvergleich für k -unabhängige Stichproben nach Kruskal-Wallis ergab einen signifikanten Unterschied bezüglich der kognitiven Leistung im DemTect (Sprachgesunde: $M = 16.10$, $SD = 1.97$; MS_SR: $M = 16.00$, $SD = 1.80$; MS_CP: $M = 12.17$, $SD = 3.31$; $\chi^2(2) = 7.21$, $p < .05$). Der anschließend durchgeführte Mann-Whitney U Test für je zwei unabhängige Stichproben ergab, dass die Gruppe der MS_CP Patienten signifikant schlechtere Leistungen im DemTect im Vergleich zu den Sprachgesunden ($U = 9.00$, $Z = -2.47$, $p < .05$) und den MS_SR Patienten ($U = 8.00$, $Z = -2.27$, $p < .05$) zeigte. Die Leistungen von Patienten mit MS_SR und Sprachgesunden waren im DemTect vergleichbar ($U = 48.50$, $Z = -.08$, $p = .94$).

Darüber hinaus können für den überwiegenden Teil der Versuchsteilnehmer (MS_SR: $n = 10$; MS_CP: $n = 5$, Sprachgesunde: $n = 10$) Daten aus dem Münchener Verständlichkeitsprofil (MVP, Ziegler & Zierdt, 2008), einem Verfahren zur Einschätzung der Sprechverständlichkeit dysarthrischer Patienten auf Basis der Wortidentifikation, präsentiert werden. Neben Aussagen zur Gesamtverständlichkeit in Prozent (%) liefert das MVP im Rahmen einer akustischen Analyse auch Angaben zum Sprechtempo und zur Sprechstimmlage (Ziegler & Zierdt, 2008). Für die vorliegende Studie sind vor allem die Ergebnisse zum Sprechtempo interessant. Die Analyse ergab ein normales Sprechtempo für die Versuchsteilnehmer aus der Gruppe der Sprachgesunden sowie ein weitestgehend normales Sprechtempo für die Patienten mit MS_SR. Bei den Patienten mit MS_CP konnte hingegen bei vier von fünf mit dem MVP untersuchten Patienten ein verlangsamtes Sprechtempo gefunden werden. Tabelle 3.1 liefert eine Übersicht über die demographischen Daten der Versuchsteilnehmer in den drei Gruppen, einschließlich der DemTect-Testergebnisse und der Einschätzung des Sprechtempos aus dem MVP.

²³Als Kriterien für die Datenerhebung galten im Rahmen dieser Studie eine möglichst gute Vergleichbarkeit der Studienteilnehmer im Hinblick auf Alter, Bildungsgrad und durchschnittliche Krankheitsdauer.

Tabelle 3.1: Übersicht über die demographischen Daten (Versuchsteilnehmer (VT), Alter, Geschlecht (m/w), Krankheitsdauer (KHD) in Monaten) sowie Test-Ergebnisse aus dem DemTect (Kalbe et al., 2004; Test-Score: 13 bis 18 = altersentsprechende Leistung; 9 bis 12 = leichte kognitive Beeinträchtigung; 0 bis 8 = Verdacht auf Demenz; k.A. = keine Angabe) und Angaben zum Sprechtempo (MVP-ST) aus dem Münchener Verständlichkeitsprofil (Ziegler & Zierdt, 2008)

VT	Gruppe	Alter	m/w	KHD	DemTect	MVP-ST
10	MS_SR	47	w	167	18	normal
11	MS_SR	45	m	182	18	normal
12	MS_SR	39	w	149	k.A.	normal
13	MS_SR	50	w	112	16	normal
14	MS_SR	38	w	31	16	normal
15	MS_SR	24	w	19	16	normal
16	MS_SR	44	w	82	18	normal
17	MS_SR	37	m	17	14	normal
18	MS_SR	37	m	17	15	normal
19	MS_SR	51	w	64	13	leicht verlangsamt
01	MS_CP	51	m	69	12	k.A.
02	MS_CP	46	m	116	13	leicht verlangsamt
03	MS_CP	66	m	106	12	verlangsamt
04	MS_CP	54	w	254	9	verlangsamt
05	MS_CP	48	m	164	18	verlangsamt
06	MS_CP	44	m	94	9	normal
20	Sprachgesund	54	w	—	18	k.A.
21	Sprachgesund	35	w	—	15	k.A.
22	Sprachgesund	59	w	—	16	k.A.
23	Sprachgesund	50	m	—	13	k.A.
24	Sprachgesund	41	w	—	15	k.A.
25	Sprachgesund	48	w	—	18	normal
26	Sprachgesund	45	m	—	k.A.	normal
27	Sprachgesund	41	w	—	k.A.	normal
28	Sprachgesund	51	w	—	k.A.	normal
29	Sprachgesund	40	w	—	18	normal
30	Sprachgesund	26	w	—	k.A.	normal
31	Sprachgesund	42	w	—	18	normal
32	Sprachgesund	37	m	—	18	normal
33	Sprachgesund	39	m	—	14	normal
34	Sprachgesund	51	w	—	k.A.	normal
35	Sprachgesund	60	m	—	14	k.A.

3.2.2 Material, Durchführung & Auswertung

Wortgenerierungsleistungen wurden in insgesamt sechs Aufgaben erhoben. Zu Beginn führten alle Versuchsteilnehmer eine lexikalische Wortgenerierungsaufgabe (LEX_A) als Übungsaufgabe durch, um mit der Aufgabenanforderung vertraut zu werden. Anschließend wurden vier Wortgenerierungsaufgaben aus dem Regensburger Wortflüssigkeitstest (RWT, Aschenbrenner et al., 2000) mit jedem Versuchsteilnehmer durchgeführt. Im Einzelnen waren das eine simple lexikalische sowie eine simple semantische Aufgabe und

zwei im Vergleich dazu komplexere Wortgenerierungsaufgaben, die einen Wechsel zwischen zwei lexikalischen bzw. semantischen Kategorien erforderten. Die Aufgabe der Versuchsteilnehmer bestand jeweils darin, innerhalb von 60 Sekunden so viele Wörter wie möglich zu produzieren.

Bei der simplen lexikalischen Aufgabe sollten die Versuchsteilnehmer Wörter mit dem Anfangsbuchstaben *M* produzieren (z.B. *Milch, mager, messen, Miete etc.* – LEX_M). Dabei sollten keine Wörter mehrfach genannt werden bzw. keine Wörter verwendet werden, die mit demselben Wortstamm beginnen. Auch das Nennen von Eigennamen war nicht zulässig. Die simple semantische Aufgabe verlangte das Generieren von Wörtern aus der Kategorie *Tiere* (z.B. *Hund, Katze, Pferd, Tiger etc.* – SEM_Tiere). Hier durften ebenfalls keine Wörter mehrfach genannt werden. Bei der lexikalischen Wechsel-Aufgabe sollten alternierend Wörter mit den Anfangsbuchstaben *H* und *T* (z.B. *Hund, Tomate, hässlich, Tante etc.* – LEX_H/T) produziert werden, während die semantische Wechsel-Aufgabe die alternierende Produktion von Wörtern aus den Kategorien *Kleidung* und *Blumen* (z.B. *Hose, Rose, Kleid, Nelke etc.* – SEM_Kleidung/Blumen) erforderte. Bei den Wechsel-Aufgaben galten darüber hinaus dieselben Einschränkungen wie bei den korrespondierenden simplen Aufgaben.

In einer weiteren Aufgabe wurden Leistungen bei der *Verbgenerierung* erhoben. Die Verbgenerierung wurde entsprechend der Anweisung in Piatt et al. (1999a; 1999b; 2004) elizitiert. Die Aufgabe der Versuchsteilnehmer bestand darin, innerhalb von 60 Sekunden, so viele "Dinge" wie möglich zu nennen, die ein Lebewesen tun kann.

Darüber hinaus wurde in einem etwa 10-minütigen Gespräch Spontansprache elizitiert, um diese in einer späteren Analyse mit den Wortgenerierungsleistungen in Beziehung zu setzen (Thiele et al., in Vorbereitung). Dabei wurden spontansprachliche Äußerungen mit den Fragen aus dem Aachener Aphasia Test (Huber et al., 1983) erhoben, wobei auf die erste Frage, die sich auf die Erkrankung bezieht, verzichtet wurde, da sie für die Elizitierung von Spontansprache bei Sprachgesunden nicht geeignet schien. Eine zusätzliche Frage bezog sich auf das Thema *Atomkraft* und fragte nach der persönlichen Meinung zu diesem Thema.

Durchführung

Die Datenerhebungen fanden in einem ruhigen Raum in einer neurologischen Rehabilitationsklinik oder bei den Versuchsteilnehmern zu Hause statt. Alle Aufgaben wurden stets in derselben Reihenfolge durchgeführt. Die Äußerungen der Versuchsteilnehmer wurden mit einem digitalen Audio-Rekorder (Zoom H4) aufgenommen und anschließend für die weitergehende Analyse in ELAN (Max Planck Institute for Psycholinguistics, 2011; Sloetjes &

Wittenburg, 2008) annotiert.

Auswertung

Die Auswertung aller Wortgenerierungsaufgaben erfolgte manuell. Für die Aufgaben LEX_M, SEM_Tiere, LEX_H/T und SEM_Kleidung/Blumen wurde die *Anzahl korrekter Nennungen* berechnet sowie eine Kategorisierung der Fehlertypen entsprechend der Anweisung im Regensburger Wortflüssigkeitstest (Aschenbrenner et al., 2000) vorgenommen. Tabelle 3.2 liefert eine Übersicht sowie Beispiele für die verschiedenen Fehlertypen. Als Perseveration wurden alle Wiederholungen zuvor genannter Wörter gewertet sowie bei den lexikalischen Aufgaben auch Wiederholungen von bedeutungsgleichen Wortstämmen. Ein Regelbruch lag vor, wenn Wörter produziert wurden, die nicht im Zusammenhang mit der vorgegebenen Aufgabenstellung standen bzw. gegen die formulierten Regeln verstießen. Bei den komplexeren Wechsel-Aufgaben wurden zwei weitere Formen von Regelbrüchen unterschieden: Kategorieperseverationen und Kategoriefehler. Ein Fehler wurde als Kategorieperseveration gewertet, wenn anstelle eines Wechsels in die zweite Kategorie, ein weiteres Wort aus der ersten Kategorie genannt wurde. Kategoriefehler waren dadurch gekennzeichnet, dass in eine andere als die beiden vorgegebenen Kategorien gewechselt wurde.

Tabelle 3.2: *Beispiele möglicher Fehlertypen in den lexikalischen und semantischen Wortgenerierungsaufgaben, Fehler sind mit * markiert*

Fehlertyp	Aufgabe	Beispiele
Perseveration	LEX_M	Mann, Maus, *Männer, Mantel ...
	SEM_Kleidung/Blumen	Hose, Rose, Hemd, *Rose ...
Einfacher Regelbruch	SEM_Tiere	Katze, Hund, Maus, *Panzer ...
	LEX_H/T	Hut, Tasche, handeln, *Toronto ...
Kategorieperseveration	LEX_H/T	Hobel, Tasche, Hemd, *Hut ...
	SEM_Kleidung/Blumen	Hose, Rose, Hemd, *Jacke ...
Kategoriefehler	LEX_HT	Hand - Tüte, Hut, *Frage ...
	SEM_Kleidung/Blumen	Hose, Rose, Jacke, *Bauch ...

In Relation zur Anzahl der Gesamtnennungen wurden für die simple lexikalische (LEX_M) und die simple semantische (SEM_Tiere) Aufgabe das Auftreten von Perseverationen und einfachen Regelbrüchen sowie für die beiden Wechsel-Aufgaben (LEX_H/T und SEM_Kleidung/Blumen) zusätzlich Kategorieperseverationen und Kategoriefehlern bestimmt. In Relation zur Gesamtzahl der Nennungen wurde dann der Anteil korrekter Nennungen sowie der Anteil der jeweiligen Fehlertypen berechnet.

Die Evaluation der Aufgabe zur Verbgenerierung erfolgte entsprechend der Vorgabe in Piatt et al. (1999b). Gemäß der gegebenen Instruktion wurden ausschließlich einmalig vorkommende Verben (ohne Ergänzung) als kor-

3 Evaluation quantitativer und qualitativer Charakteristika der Wortgenerierungsleistung bei Multipler Sklerose

rekte Nennung gewertet, z.B. *sammeln, fliegen, sprechen, essen, brüten, sehen, rufen, schreien ...*. Perseverationen, Mehrwortäußerungen, Umschreibungen sowie Wörter aus anderen Wortklassen wurden als fehlerhafte Nennung kodiert, wobei die drei letztgenannten unter dem Terminus *Regelbruch* zusammengefasst wurden. Beispiele für Perseverationen und Regelbrüche in der Verbgenerierungsaufgabe finden sich in der folgenden Tabelle (Tabelle 3.3):

Tabelle 3.3: *Kategorisierung der Fehlertypen in der Verbgenerierungsaufgabe, Fehler sind mit * markiert*

Fehlertyp	Beispiel
Perseveration	fliegen, tauchen, singen, *tauchen ...
Regelbruch	gehen, putzen, trinken, *Kaffee kochen ... schreiben, Kaffee kochen, gehen, *Suppe kochen ... schreiben, putzen, *aufs Klo gehen, essen ... singen, essen, gehen, *fröhlich sein ...

Die Auswertung der Spontansprache erfolgte ebenfalls manuell und fokussierte insbesondere lexikalisch-semantische Charakteristika (Wortfindungsstörungen, lexikalische Dichte etc.), Aspekte syntaktischer Komplexität (mittlere Phrasenlänge etc.) sowie ein Maß zur Beurteilung der Formelhaftigkeit spontansprachlicher Äußerungen (Thiele et al., in Vorbereitung). Die vorliegende Arbeit konzentriert sich auf die Analyse der Wortgenerierungsdaten.

Sämtliche statistischen Analysen wurden mit der Statistiksoftware PASW 18 (2009) durchgeführt. Bei jedem Vergleich der Gruppenunterschiede einzelner Variablen wurde zunächst mit einem nicht-parametrischen Test für k -unabhängige Stichproben nach Kruskal-Wallis überprüft, ob sich signifikante Unterschiede bezüglich der untersuchten Variablen zwischen den drei Gruppen (Sprachgesund vs. MS_SR vs. MS_CP) fanden. Zeigte der Test einen signifikanten Unterschied zwischen den drei Gruppen ($p < .05$), wurden anschließend mit dem Mann-Whitney U Test Unterschiede zwischen den einzelnen Gruppen (Sprachgesund vs. MS_SR, Sprachgesund vs. MS_CP und MS_SR vs. MS_CP) für die jeweilige Variable überprüft. Auf nicht-parametrische Verfahren wurde aufgrund der geringen Stichprobengrößen sowie der deutlichen Unterschiede in den Stichprobengrößen zwischen den drei Gruppen zurückgegriffen.

3.3 Ergebnisse

Im Folgenden werden zunächst die Ergebnisse der Wortgenerierungsleistungen in den lexikalischen und semantischen Aufgaben sowie bei der Verbgenerierung dargestellt, bevor anschließend die Analyse des Anteils korrekter

Nennungen sowie der Fehlertypen folgt.

3.3.1 Wortgenerierungsleistungen

Um zu ermitteln, ob sich die Wortgenerierungsleistung, gemessen in der *Anzahl korrekter Nennungen*, generell zwischen den Sprachgesunden und den Patienten mit MS_SR und MS_CP unterschied wurde zunächst ein Wortgenerierungs-gesamtwert berechnet. Dieser ergab sich aus der *Anzahl korrekter Nennungen* gemittelt über die vier Wortgenerierungsaufgaben aus dem RWT (Aschenbrenner et al., 2000). Anhand dieses Wertes wurde überprüft, ob sich die Wortgenerierungsleistung insgesamt zwischen den drei Gruppen unterschied. Die Berechnung ergab einen signifikanten Unterschied zwischen den drei Gruppen für den Wortgenerierungs-gesamtwert (Sprachgesunde: $M = 17.00$, $SD = 2.90$; MS_SR: $M = 12.88$, $SD = 2.30$; MS_CP: $M = 8.51$, $SD = 3.78$; $\chi^2(2) = 18.28$, $p < .001$). Die anschließende Berechnung gruppenspezifischer Unterschiede ergab signifikante Unterschiede zwischen der Gruppe der Sprachgesunden und der Gruppe der Patienten mit MS_SR ($U = 18.00$, $Z = -3.27$, $p < .01$) sowie zwischen der Gruppe der Sprachgesunden und der Gruppe der Patienten mit MS_CP ($U = 2.00$, $Z = -3.40$, $p < .001$). Im Vergleich zu den Sprachgesunden produzierten sowohl die Patienten mit MS_SR als auch die Patienten mit MS_CP im Mittel weniger korrekte Nennungen. Das heißt, beide Patientengruppen zeigten im Vergleich zu den Sprachgesunden insgesamt schlechtere Wortgenerierungsleistungen. Der Vergleich der beiden Patientengruppen untereinander ergab lediglich Hinweise auf einen tendenziellen Unterschied ($U = 12.00$, $Z = -1.96$, $p = .50$) in Richtung einer besseren Leistung für die Gruppe der Patienten mit MS_SR im Vergleich zu den Patienten mit MS_CP.

Spezifische Wortgenerierungsleistungen in semantischen und lexikalischen Aufgaben

Es wurde überprüft, ob sich die drei Gruppen signifikant hinsichtlich ihrer Leistung in den verschiedenen Wortgenerierungsaufgaben unterschieden. Die Analyse ergab für die Aufgabe LEX_M, dass sich Sprachgesunde, Patienten mit MS_SR und Patienten mit MS_CP bezüglich der Wortgenerierungsleistung (Sprachgesunde: $M = 11.06$, $SD = 3.75$; MS_SR: $M = 8.80$, $SD = 2.49$; MS_CP: $M = 7.17$, $SD = 3.00$; $\chi^2(2) = 5.64$, $p = .06$) lediglich tendenziell unterschieden (siehe auch Abbildung 3.1). Auf eine weitergehende Analyse mit dem Mann-Whitney U Test wurde daher verzichtet.

Für die Aufgabe SEM_Tiere zeigte sich ein signifikanter Unterschied in der Wortgenerierungsleistung zwischen den Gruppen (Sprachgesunde: $M = 26.06$, $SD = 4.02$; MS_SR: $M = 20.50$, $SD = 4.81$; MS_CP: $M = 14.60$, $SD = 5.37$; $\chi^2(2) = 14.10$, $p < .01$; siehe auch Abbildung 3.2). Auch für die beiden Wechsel-Aufgaben LEX_H/T (Sprachgesunde: $M = 15.44$, $SD =$

3 Evaluation quantitativer und qualitativer Charakteristika der Wortgenerierungsleistung bei Multipler Sklerose

4.47; MS_SR: $M = 10.30$, $SD = 2.98$; MS_CP: $M = 6.50$, $SD = 3.78$; $\chi^2(2) = 16.24$, $p < .001$, siehe auch Abbildung 3.3) und SEM_Kleidung/Blumen (Sprachgesunde: $M = 15.44$, $SD = 2.85$; MS_SR: $M = 11.90$, $SD = 3.99$; MS_CP: $M = 7.50$, $SD = 4.64$; $\chi^2(2) = 14.79$, $p < .01$, siehe auch Abbildung 3.4) konnten signifikante Gruppenunterschiede gefunden werden.

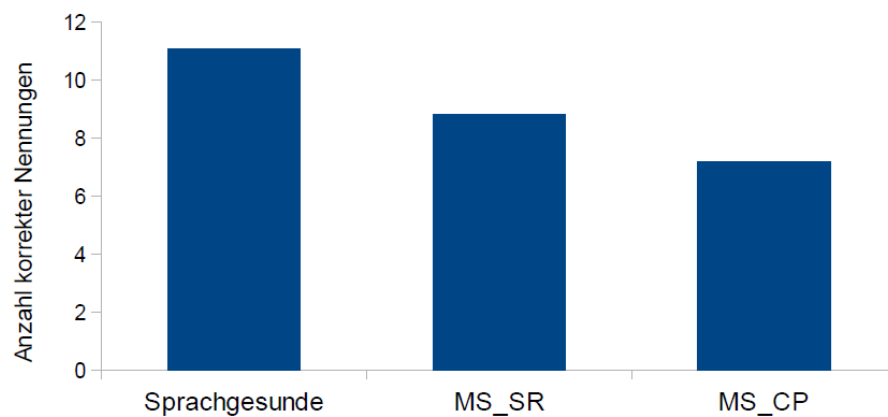


Abbildung 3.1: Aufgabe LEX_M: Anzahl korrekter Nennungen

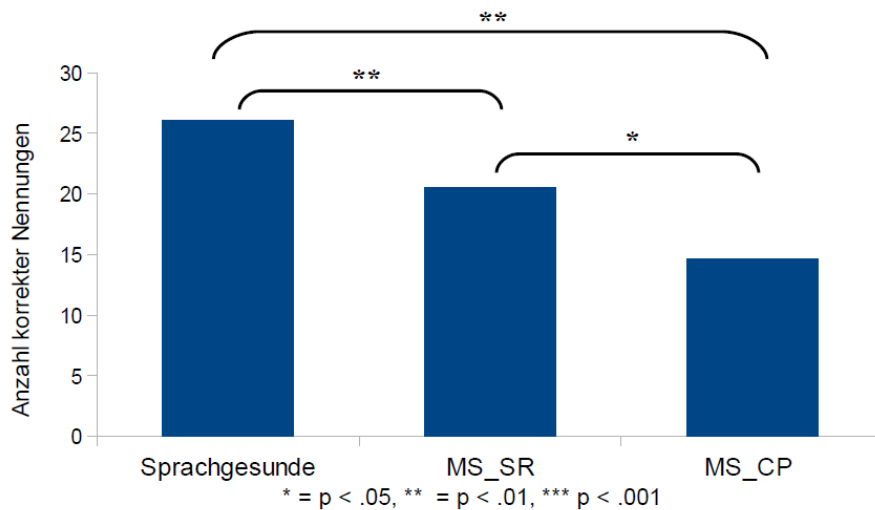


Abbildung 3.2: Aufgabe SEM_Tiere: Anzahl korrekter Nennungen

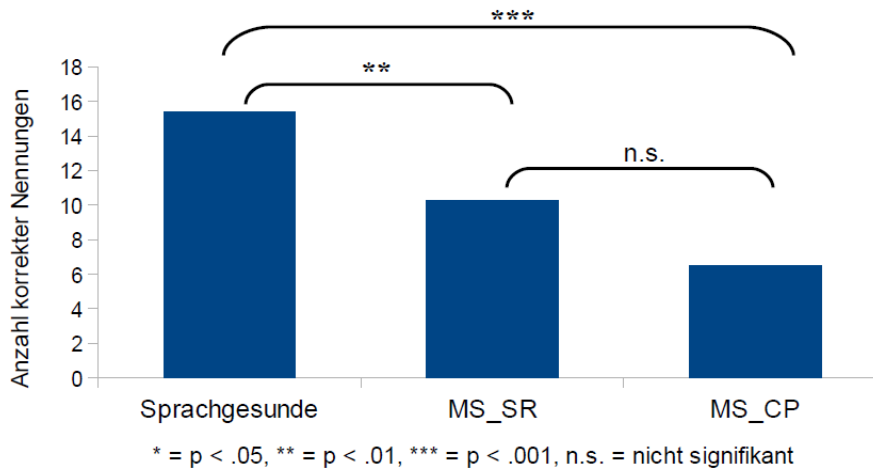


Abbildung 3.3: Aufgabe LEX_H/T: Anzahl korrekter Nennungen

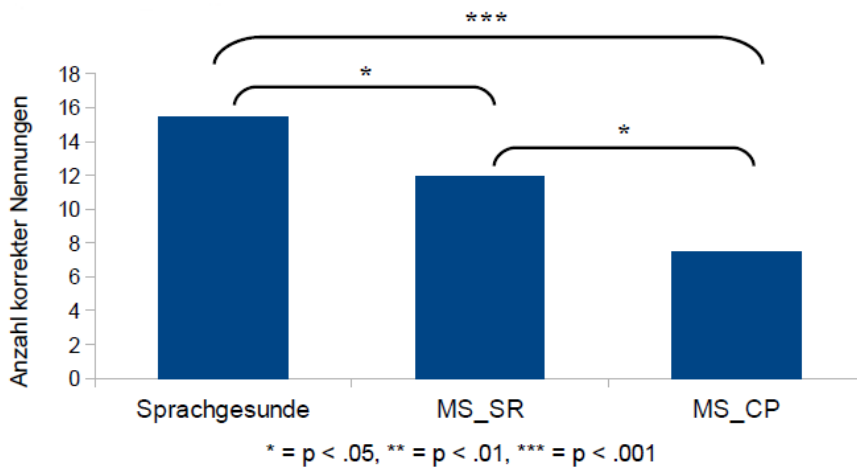


Abbildung 3.4: Aufgabe SEM_Kleidung/Blumen: Anzahl korrekter Nennungen

Die anschließende Analyse der Wortgenerierungsleistungen zwischen einzelnen Gruppen ergab für die Aufgabe SEM_Tiere signifikante Unterschiede zwischen allen Gruppen (Sprachgesunde vs. MS_SR: $U = 26.50$, $Z = -2.83$, $p < .01$; Sprachgesunde vs. MS_CP: $U = 5.00$, $Z = -2.90$, $p < .01$ und MS_SR vs. MS_CP: $U = 8.50$, $Z = -2.04$, $p < .05$; siehe auch Abbildung 3.2). Die Gruppe der Sprachgesunden produzierte signifikant mehr korrekte Nennungen als die beiden Gruppen der Multiple Sklerose Patienten. Die Patienten MS_SR produzierten mehr korrekte Nennungen im Vergleich zu den Patienten mit MS_CP. Das heißt, die Wortgenerierungsleistung der Sprachgesunden war in der Aufgabe SEM_Tiere im Vergleich zu beiden Patientengruppen besser. Patienten mit MS_SR zeigten wiederum bessere Wortgenerierungsleistungen in der simplen semantischen Aufgabe als Patienten mit MS_CP.

Bei den Wechsel-Aufgaben konnten für die Aufgabe LEX_H/T signifikante Unterschiede bezüglich der Wortgenerierungsleistung zwischen den Sprachgesunden und den Patienten mit MS_SR ($U = 26.00$, $Z = -2.86$, $p < .01$) sowie zwischen den Sprachgesunden und den Patienten mit MS_CP ($U = 2.50$, $Z = -3.37$, $p < .001$) gefunden werden. Dabei produzierten die sprachgesunden Versuchsteilnehmer jeweils signifikant mehr korrekte Nennungen im Vergleich zu den Patienten mit Multipler Sklerose, was einer besseren Wortgenerierungsleistung entspricht. Die beiden Patientengruppen unterschieden sich bezüglich ihrer Wortgenerierungsleistungen nur tendenziell ($U = 12.00$, $Z = -1.96$, $p = .50$) in Richtung von numerisch mehr korrekten Nennungen für die Patienten mit MS_SR im Vergleich zur Gruppe der Patienten mit MS_CP (siehe auch Abbildung 3.3).

Bei der semantischen Wechsel-Aufgabe SEM_Kleidung/Blumen zeigten sich signifikante Unterschiede in den Wortgenerierungsleistungen zwischen Sprachgesunden und Patienten mit MS_SR ($U = 32.00$, $Z = -2.55$, $p < .05$), zwischen Sprachgesunden und Patienten mit MS_CP ($U = 3.50$, $Z = -3.30$, $p < .001$) sowie zwischen den Patienten mit MS_SR und denen mit MS_CP ($U = 11.00$, $Z = -2.07$, $p < .05$). Dabei waren bessere Wortgenerierungsleistung für die Gruppe der Sprachgesunden gegenüber den beiden Patientengruppen, sowie eine bessere Leistung für die Patienten mit MS_SR im Vergleich zur Gruppe der Patienten mit MS_CP zu erkennen (siehe auch Abbildung 3.4).

Verbgenerierung

Der nicht-parametrische Test für k -unabhängige Stichproben nach Kruskal-Wallis ergab signifikante Unterschiede zwischen den drei Gruppen bezüglich der *Anzahl korrekter Nennungen* bei der Verbgenerierung (Sprachgesunde: $M = 18.31$, $SD = 6.00$; MS_SR: $M = 16.00$, $SD = 3.68$; MS_CP.: $M = 9.83$, $SD = 4.96$; $\chi^2(2) = 8.02$, $p < .05$). Eine anschließende Analyse der

Unterschiede zwischen einzelnen Gruppen zeigte signifikante Unterschiede zwischen den Sprachgesunden und den Patienten mit MS_CP ($U = 12.5$, $Z = -2.63$, $p < .01$) sowie zwischen den Patienten mit MS_SR und denen mit MS_CP ($U = 9.00$, $Z = -2.30$, $p < .05$). Patienten mit MS_CP produzierten jeweils signifikant weniger korrekte Nennungen als die jeweilige Vergleichsgruppe, was einer signifikant schlechteren Verbgenerierungsleistung entsprach. Die Gruppe der Patienten mit MS_SR zeigte im Vergleich zu den Sprachgesunden ähnliche Leistungen bei der Verbgenerierung ($U = 64.00$, $Z = -0.85$, $p = .40$). Die folgende Abbildung 3.5 veranschaulicht die Ergebnisse.

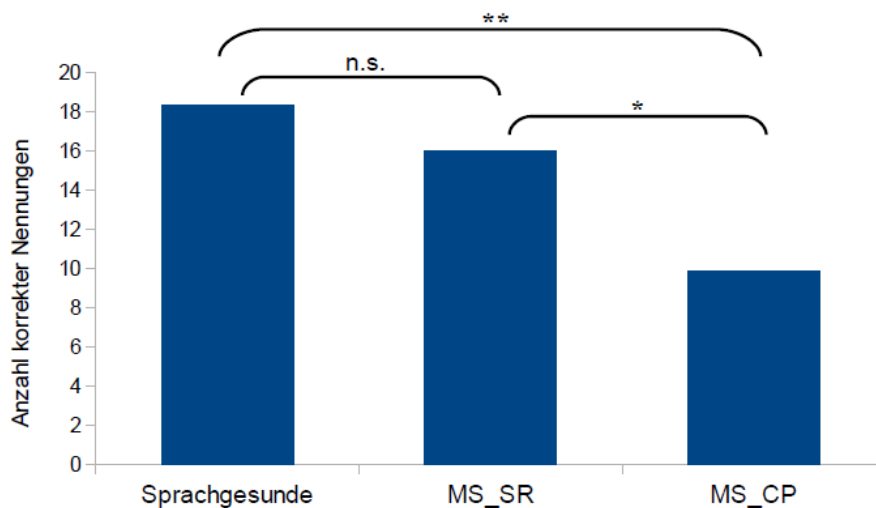


Abbildung 3.5: Verbgenerierung: Anzahl korrekter Nennungen

Zusammengefasst fiel beim Vergleich der drei Gruppen in den lexikalischen und semantischen Wortgenerierungsaufgaben sowie bei der Aufgabe zur Verbgenerierung ein treppenähnliches Leistungsprofil auf. Die *Anzahl korrekter Nennungen* war dabei numerisch für die Gruppe der Sprachgesunden immer am höchsten, gefolgt von den Patienten mit MS_SR. Die numerisch geringste *Anzahl korrekter Nennungen* zeigten die Patienten aus der MS_CP Gruppe. Für alle Wortgenerierungsaufgaben (außer LEX_M) konnten signifikante Gruppenunterschiede bezüglich der Wortgenerierungsleistung gefunden werden. In den semantischen Aufgaben (SEM_Tiere und SEM_Kleidung/Blumen) waren bessere Leistungen für die Gruppe der Sprachgesunden gegenüber beiden Patientengruppen sowie bessere Leistungen für Patienten mit MS_SR im Vergleich zu Patienten mit MS_CP zu beobachten. Bei der lexikalischen Wechsel-Aufgabe zeigten sich reliable Leistungsunterschiede nur im Vergleich der Sprachgesunden gegenüber den beiden Patientengruppen. Dabei waren die Leistungen Sprachgesunder im Ver-

gleich zu den Patienten mit Multipler Sklerose signifikant besser. Patienten mit MS_SR und MS_CP unterschieden sich bezüglich ihrer Wortgenerierungsleistungen nicht in dieser Aufgabe. Bei der Verbgenerierung zeigten die Gruppen der Sprachgesunden und der Patienten mit MS_SR bessere Leistungen im Vergleich zur Gruppe der Patienten mit MS_CP. Patienten mit MS_SR unterschieden sich bezüglich der Leistungen bei der Verbgenerierung nicht von den Sprachgesunden.

3.3.2 Anteil korrekter Nennungen & Evaluation der Fehler-typen

Darüber hinaus war von Interesse, ob sich der Anteil korrekter Nennungen in Relation zur Gesamtzahl der Nennungen zwischen den Gruppen unterschied. Dazu wurde der prozentuale Anteil korrekter Nennungen, als indirektes Fehler-Maß, für die einzelnen Wortgenerierungsaufgaben berechnet:

$$\text{Nennungen_korrekt}/\text{Nennungen_gesamt} \times 100$$

Je höher der Anteil korrekter Nennungen, desto weniger falsche Nennungen wurden produziert, das heißt, umso besser war die jeweilige Leistung.

Lexikalische und semantische Wortgenerierung

Zunächst wurde ermittelt, ob sich die drei Gruppen generell bezüglich des Anteils korrekter Nennungen unterschieden. Hierbei zeigte sich nur für die lexikalische Wechsel-Aufgabe (LEX_H/T) ein signifikanter Unterschied bezüglich des Anteils korrekter Nennungen zwischen den Gruppen (Sprachgesunde: $M = 95.55$, $SD = 5.93$; MS_SR: $M = 95.48$, $SD = 4.90$; MS_CP: $M = 62.79$, $SD = 35.20$; $\chi^2(2) = 8.01$, $p < .05$). Für die Aufgaben LEX_M (Sprachgesunde: $M = 91.90$, $SD = 8.65$; MS_SR: $M = 87.10$, $SD = 13.42$; MS_CP: $M = 88.17$, $SD = 9.39$; $\chi^2(2) = .93$, $p = .63$), SEM_Tiere (Sprachgesunde: $M = 98.54$, $SD = 1.96$; MS_SR: $M = 98.56$, $SD = 2.46$; MS_CP: $M = 95.99$, $SD = 4.01$; $\chi^2(2) = 2.66$, $p = .26$) und SEM_Kleidung/Blumen (Sprachgesunde: $M = 91.12$, $SD = 12.06$; MS_SR: $M = 86.79$, $SD = 23.26$; MS_CP: $M = 73.64$, $SD = 33.48$; $\chi^2(2) = 2.98$, $p = .23$) konnten keine signifikanten Unterschiede bezüglich des Anteils korrekter Nennungen zwischen den Gruppen gefunden werden. Die Abbildungen 3.6, 3.7, 3.8 und 3.9 veranschaulichen diese Ergebnisse²⁴.

²⁴Für eine übersichtliche Darbietung und aus Gründen der besseren Vergleichbarkeit wurde bei der Darstellung der Ergebnisse auf Balkendiagramme zurückgegriffen und auf die üblicherweise verwendeten Kreisdiagramme verzichtet.

Ein anschließender Vergleich zwischen einzelnen Gruppen für die Aufgabe LEX_H/T ergab signifikante Unterschiede bezüglich des Anteils korrekter Nennungen zwischen den Sprachgesunden und den Patienten mit MS_CP ($U = 15.00$, $Z = -2.56$, $p < .05$) sowie zwischen den Patienten mit MS_SR und denen mit MS_CP ($U = 7.50$, $Z = -2.51$, $p < .05$). Der Anteil korrekter Nennungen unterschied sich hingegen nicht zwischen den Sprachgesunden und den Patienten mit MS_SR ($U = 76.00$, $Z = -0.23$, $p = .86$). Während der Anteil korrekter Nennungen für die Patienten mit MS_CP bei etwa 65% lag, zeigten die Sprachgesunden sowie die Patienten mit MS_SR mit einem Anteil korrekter Nennungen in Höhe von 95% signifikant bessere Leistungen (siehe Abbildung 3.8).

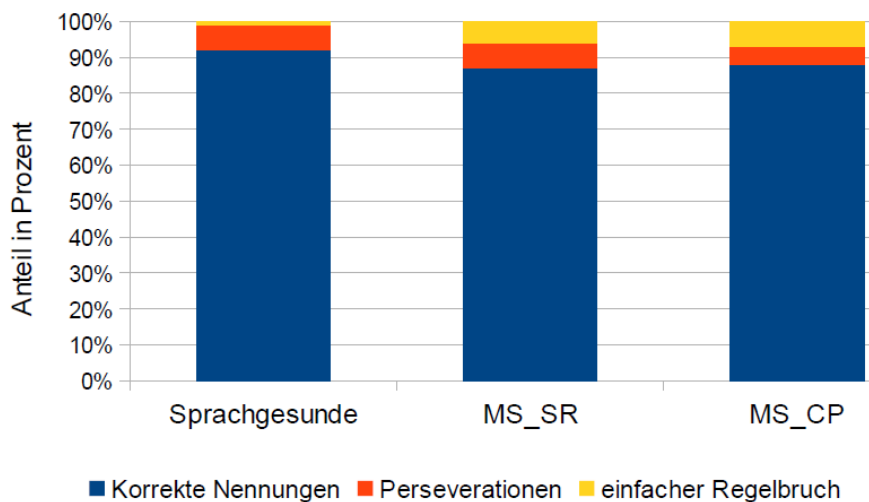


Abbildung 3.6: Aufgabe LEX_M: Anteil korrekter Nennungen, Perseverationen und einfacher Regelbrüche in allen Gruppen

Die Analyse des indirekten Fehlermaßes konnte nur bedingt Unterschiede zwischen den drei Gruppen aufdecken. Der Anteil korrekter Nennungen unterschied sich lediglich in der Aufgabe LEX_H/T zwischen den Gruppen, mit einem signifikant größeren Anteil korrekter Nennungen für die Sprachgesunden im Vergleich zu den Patienten mit MS_CP sowie im Vergleich von Patienten mit MS_SR zu Patienten mit MS_CP. Die folgende Analyse beschäftigt sich nun mit der Frage nach Gruppenunterschieden bezüglich der auftretenden Fehlertypen. Für die einfache lexikalische und die simple semantische Aufgabe wurde die Auftretenshäufigkeit bzw. der Anteil von Perseverationen und einfachen Regelbrüchen in Relation zur Gesamtzahl der Nennungen zwischen den Gruppen miteinander verglichen.

3 Evaluation quantitativer und qualitativer Charakteristika der Wortgenerierungsleistung bei Multipler Sklerose

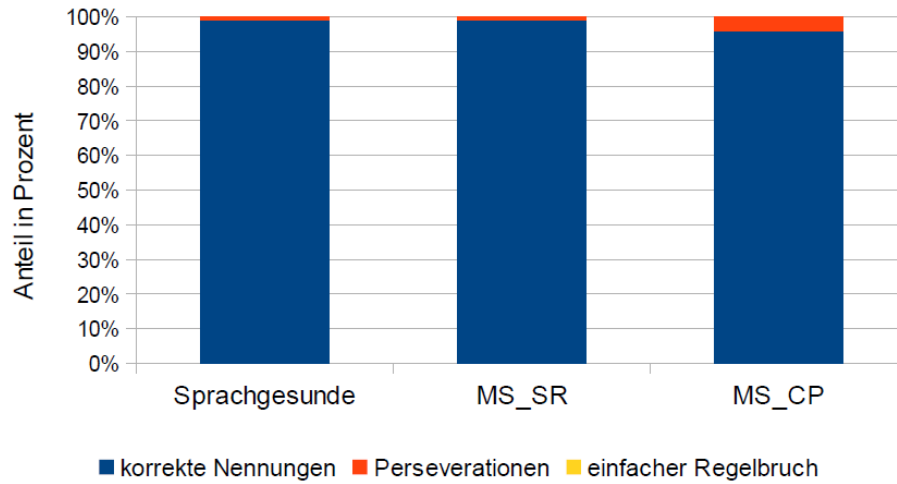


Abbildung 3.7: Aufgabe SEM_Tiere: Anteil korrekter Nennungen, Perseverationen und einfacher Regelbrüche in allen Gruppen

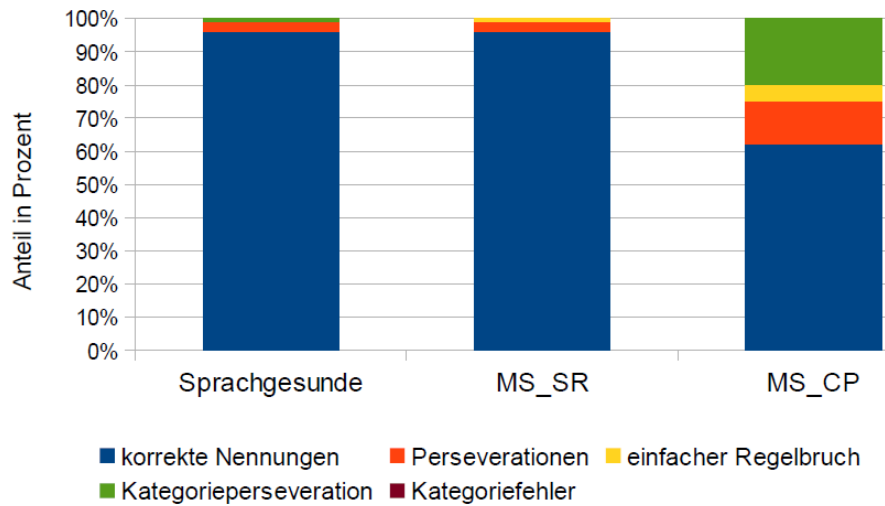


Abbildung 3.8: Aufgabe LEX_H/T: Anteil korrekter Nennungen, Perseverationen, einfacher Regelbrüche, Kategorieperseverationen und Kategoriefehler in allen Gruppen

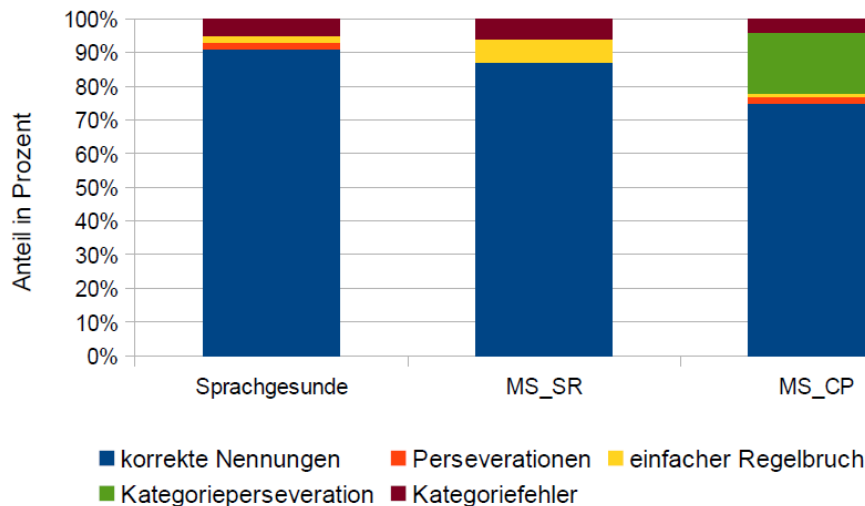


Abbildung 3.9: Aufgabe SEM_Kleidung/Blumen: Anteil korrekter Nennungen, Perseverationen, einfacher Regelbrüche, Kategorieperseverationen und Kategoriefehler in allen Gruppen

Die Berechnung eines generellen Unterschieds zwischen den drei Gruppen für die Aufgabe LEX_M ergab keinen signifikanten Unterschied bezüglich des Anteils von Perseverationen (Sprachgesunde: $M = 7.41$, $SD = 8.84$; MS_SR: $M = 6.78$, $SD = 7.76$; MS_CP: $M = 4.99$, $SD = 5.64$; $\chi^2(2) = 0.29$, $p = .86$) und einfachen Regelbrüchen (Sprachgesunde: $M = 0.69$, $SD = 2.78$; MS_SR: $M = 6.12$, $SD = 13.46$, MS_CP: $M = 6.85$, $SD = 11.76$; $\chi^2(2) = 3.3$, $p = .19$; siehe auch Abbildung 3.6). Für die simple semantische Aufgabe SEM_Tiere konnte ebenfalls kein Unterschied bezüglich des Anteils der Perseverationen (Sprachgesunde: $M = 1.18$, $SD = 1.81$; MS_SR: $M = 1.44$, $SD = 2.46$; MS_CP: $M = 4.01$, $SD = 4.01$; $\chi^2(2) = 3.22$, $p = .20$) und der einfachen Regelbrüche (Sprachgesunde: $M = 0.28$, $SD = 1.14$; MS_SR: $M = 0.00$, $SD = 0.00$; MS_CP: $M = 0.00$, $SD = 0.00$; $\chi^2(2) = 0.94$, $p = .63$) zwischen den drei Gruppen gefunden werden (vgl. Abbildung 3.7).

Für die Wechsel-Aufgaben wurde zwischen weiteren Fehlertypen unterschieden. Gruppenunterschiede wurden für den Anteil von Perseverationen, einfachen Regelbrüchen, Kategorieperseverationen und Kategoriefehlern berechnet.

Die Überprüfung auf das Vorliegen eines generellen Unterschieds ergab für die Aufgabe LEX_H/T lediglich bezüglich des Anteils einfacher Regelbrüche einen signifikanten Unterschied zwischen den drei Gruppen (Sprach-

gesunde: $M = 0.42$, $SD = 1.67$; MS_SR: $M = 1.11$, $SD = 3.51$; MS_CP: $M = 5.33$, $SD = 6.21$; $\chi^2(2) = 6.86$, $p < .05$). Der Anteil von Perseverationen (Sprachgesunde: $M = 3.14$, $SD = 4.03$; MS_SR: $M = 3.41$, $SD = 4.48$; MS_CP: $M = 12.60$, $SD = 12.14$; $\chi^2(2) = 4.20$, $p = .12$), Kategorieperseverationen (Sprachgesunde: $M = 0.90$, $SD = 2.46$; MS_SR: $M = 0.00$, $SD = 0.00$; MS_CP: $M = 19.29$, $SD = 34.69$; $\chi^2(2) = 4.16$, $p = .13$) und Kategoriefehlern (Sprachgesunde: $M = 0.00$, $SD = 0.00$; MS_SR: $M = 0.00$, $SD = 0.00$; MS_CP: $M = 0.00$, $SD = 0.00$; $\chi^2(2) = 0.00$, $p = 1.00$) unterschied sich in der lexikalischen Wechsel-Aufgabe nicht zwischen den drei Gruppen. Ein anschließender Test auf signifikante Gruppenunterschiede zwischen einzelnen Gruppen für die einfachen Regelbrüche ergab lediglich zwischen der Gruppe der Sprachgesunden und den Patienten mit MS_CP einen signifikanten Unterschied mit einem signifikant größerem Anteil von Regelbrüchen für die Patienten mit MS_CP im Vergleich zu den Sprachgesunden ($U = 25.50$, $Z = -2.47$, $p < .05$). Der Anteil einfacher Regelbrüche unterschied sich hingegen nicht zwischen den Sprachgesunden und den Patienten mit MS_SR ($U = 76.5$, $Z = -0.40$, $p = .69$) bzw. zwischen den Patienten mit MS_SR und denen mit MS_CP ($U = 18.5$, $Z = -1.64$, $p = .10$).

Für die semantische Wechsel-Aufgabe zeigten sich Hinweise auf einen signifikanten Unterschied zwischen den Gruppen nur für den Anteil der Kategorieperseverationen (Sprachgesunde: $M = 0.00$, $SD = 0.00$; MS_SR: $M = 0.00$, $SD = 0.00$; MS_CP: $M = 18.27$, $SD = 34.01$; $\chi^2(2) = 8.95$, $p < .05$). Der Anteil der Perseverationen (Sprachgesunde: $M = 2.37$, $SD = 2.80$; MS_SR: $M = 0.00$, $SD = 0.00$; MS_CP: $M = 2.39$, $SD = 3.72$; $\chi^2(2) = 5.32$, $p = .07$), der einfachen Regelbrüche (Sprachgesunde: $M = 1.83$, $SD = 4.86$; MS_SR: $M = 6.96$, $SD = 15.71$; MS_CP: $M = 1.39$, $SD = 3.40$; $\chi^2(2) = 0.81$, $p = .67$) und der Kategoriefehler (Sprachgesunde: $M = 4.68$, $SD = 10.58$; MS_SR: $M = 6.25$, $SD = 15.87$; MS_CP: $M = 4.31$, $SD = 6.68$; $\chi^2(2) = .249$, $p = .88$) unterschied sich hingegen nicht zwischen den drei Gruppen. Ein anschließender Vergleich zwischen einzelnen Gruppen für den Anteil der Kategorieperseverationen ergab einen signifikanten Unterschied zwischen den Sprachgesunden und den Patienten mit MS_CP ($U = 32.00$, $Z = -2.36$, $p < .05$) sowie einen tendenziellen Unterschied zwischen den beiden Patientengruppen ($U = 20.00$, $Z = -1.89$, $p = .06$), jeweils in Richtung eines größeren Anteils von Kategorieperseverationen für die Patienten mit MS_CP im Vergleich zu den Sprachgesunden bzw. gegenüber den Patienten mit MS_SR. Zwischen den Sprachgesunden und den Patienten mit MS_SR ($U = 80.00$, $Z = 0.00$, $p = 1.00$) konnte kein Unterschied bzgl. des Anteils der Kategorieperseverationen gefunden werden.

Verbgenerierung

Auch bei der Verbgenerierung wurde untersucht, ob sich der Anteil korrekter Nennungen in Relation zur Gesamtzahl der Nennungen (als indirektes Fehlermaß) sowie der Anteil der Perseverationen und Regelbrüche zwischen den Gruppen unterschied. Abbildung 3.10 veranschaulicht die jeweiligen Anteile in den drei Gruppen. Der Test nach Kruskal-Wallis ergab keinen Unterschied bezüglich des Anteils korrekter Nennungen (Sprachgesunde: $M = 82.96$, $SD = 14.79$; MS_SR: $M = 89.16$, $SD = 9.84$; MS_CP: $M = 77.40$, $SD = 21.84$; $\chi^2(2) = 1.11$, $p = .58$) sowie bezüglich des Anteils der Perseverationen (Sprachgesunde: $M = 3.26$, $SD = 5.08$; MS_SR: $M = 3.99$, $SD = 4.89$; MS_CP: $M = 6.37$, $SD = 5.54$; $\chi^2(2) = 2.16$, $p = .34$) und der Regelbrüche (Sprachgesunde: $M = 13.78$, $SD = 15.57$; MS_SR: $M = 6.84$, $SD = 8.30$; MS_CP: $M = 16.23$, $SD = 23.56$; $\chi^2(2) = 1.02$, $p = .60$) zwischen den Gruppen.

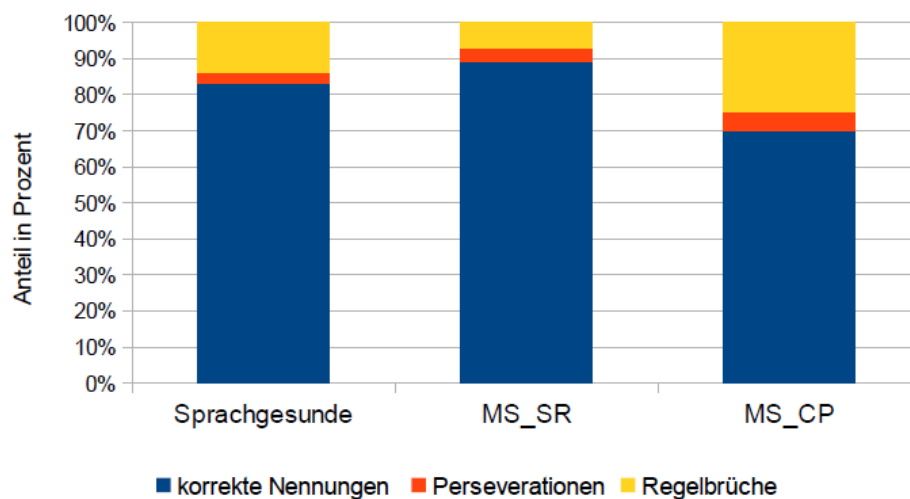


Abbildung 3.10: Verbgenerierung: Anteil korrekter Nennungen, Perseverationen und Regelbrüche in allen Gruppen

Zusammengenommen zeigten sich Unterschiede bezüglich des Fehleranteils zwischen den Gruppen lediglich für die beiden komplexeren Wechsel-Aufgaben. Bei der lexikalischen Wechsel-Aufgabe unterschied sich der Anteil einfacher Regelbrüche zwischen den Gruppen mit signifikant weniger Regelbrüchen für die Sprachgesunden im Vergleich zu den Patienten mit MS_CP. In der semantischen Wechsel-Aufgabe zeigte sich ein signifikant geringerer Anteil von Kategorieperseverationen bei den Sprachgesunden im Vergleich zu den Patienten mit MS_CP sowie ein tendenziell geringerer Anteil von Kategorieperseverationen bei den Patienten mit MS_SR im Vergleich zu Patienten mit MS_CP.

3.4 Diskussion

Gegenstand der vorliegenden Untersuchung war die Evaluation von Wortgenerierungsleistungen in lexikalischen, semantischen und Verbgenerierungsaufgaben bei Patienten mit MS_SR, Patienten mit MS_CP sowie bei sprachgesunden Probanden. Ziel der Studie war, durch eine Kombination von quantitativen und qualitativen Maßen einer genaueren Charakterisierung der Wortgenerierungskompetenz von Patienten mit Multipler Sklerose näher zu kommen.

Die Analyse des Wortgenerierungsgesamtwertes ergab Hinweise darauf, dass sich die Wortgenerierungsfähigkeit der Patienten mit Multipler Sklerose beider Verlaufsformen von den Leistungen Sprachgesunder unterscheiden. Die Evaluation spezifischer Wortgenerierungsleistungen in den einzelnen Aufgaben ergab für die simple lexikalische Aufgabe (LEX_M) lediglich einen tendenziellen Unterschied zwischen den drei Gruppen. Signifikant bessere Wortgenerierungsleistungen für die Gruppe der Sprachgesunden im Vergleich zu den Multiple Sklerose Patienten zeigten sich unabhängig von der Verlaufsform der Erkrankung in der simplen semantischen (SEM_Tiere) sowie in der lexikalischen (LEX_H/T) und der semantischen (SEM_Kleidung/Blumen) Wechsel-Aufgabe. Die beiden semantischen Aufgaben konnten darüber hinaus zwischen den beiden Verlaufsformen der Multiplen Sklerose dissoziieren. Bei der Verbgenerierung zeigten Patienten mit MS_CP signifikant schlechtere Leistungen im Vergleich zu den Sprachgesunden und zu den Patienten mit schubförmig-remittierendem Verlauf der Erkrankung.

Mit der Analyse des indirekten Fehlermaßes konnten lediglich für die lexikalische Wechsel-Aufgabe signifikante Unterschiede bezüglich des Anteils korrekter Nennungen zwischen den Gruppen gefunden werden. Dabei produzierte die Gruppe der Sprachgesunden sowie die Gruppe der Patienten mit MS_SR anteilig mehr korrekte Nennungen im Vergleich zu den Patienten mit MS_CP. Die anschließende Evaluation des Anteils spezifischer Fehlertypen ergab nur für die beiden Wechsel-Aufgaben Hinweise auf einen Unterschied zwischen Sprachgesunden, Patienten mit MS_SR und Patienten mit MS_CP. In der lexikalischen Wechsel-Aufgabe produzierten Patienten mit MS_CP signifikant mehr einfache Regelbrüche im Vergleich zu den Sprachgesunden. In der semantischen Wechsel-Aufgabe fanden sich signifikant weniger Kategorieperseverationen für die Gruppe der Sprachgesunden im Vergleich zu den Patienten mit MS_CP sowie eine Tendenz in Richtung eines Unterschied mit weniger Kategorieperseverationen für die Gruppe der Patienten mit MS_SR im Vergleich zur Gruppe der Patienten mit MS_CP. Bei der Evaluation der Verbgenerierungsleistungen fanden sich weder Unterschiede bezüglich des Anteils korrekter Nennungen noch in Bezug auf den Anteil der Perseverationen und Regelbrüche zwischen den Gruppen.

3.4.1 Interpretation der Wortgenerierungsleistungen

Die Ergebnisse weisen darauf hin, dass sowohl Patienten mit MS_SR als auch Patienten mit MS_CP im Vergleich zu Sprachgesunden in ihren Wortgenerierungsleistungen beeinträchtigt sind. Dies ist kein überraschendes Ergebnis, da die Evaluation von Wortgenerierungsleistungen als ein sensitives Maß zur Identifizierung kognitiver Defizite bei Multipler Sklerose beschrieben wird (Beatty, 2002; Henry & Beatty, 2006; Miller & Leary, 2007). Leistungsunterschiede zwischen Patienten mit MS_SR und MS_CP fanden sich bei einem Vergleich der allgemeinen Wortgenerierungskompetenz jedoch nicht. Das lässt darauf schließen, dass, sollten Unterschiede bestehen, sich diese eher bei der Evaluation spezifischer Wortgenerierungsleistungen in Form quantitativer Unterschiede zu erkennen geben oder auf qualitative Unterschiede in den Leistungen zurückzuführen sind.

Unterschiede zwischen Sprachgesunden und Patienten mit Multipler Sklerose in spezifischen Wortgenerierungsleistungen fanden sich in den beiden semantischen Aufgaben sowie in der lexikalischen Wechsel-Aufgabe. Diese drei Aufgaben scheinen also gut zwischen den Leistungen Sprachgesunder und denen von Patienten mit Multipler Sklerose dissoziieren zu können. Für die simple lexikalische Aufgabe fanden sich Hinweise auf einen tendenziellen Unterschied der Wortgenerierungsleistungen zwischen den Gruppen. Weitere Studien mit größeren Stichproben sind hier notwendig, um verlässlichere Aussagen machen zu können. Es ist denkbar, dass sich auch bei simplen lexikalischen Wortgenerierungsaufgaben reliable Unterschiede zwischen Sprachgesunden und Patienten mit Multipler Sklerose finden lassen (Drake et al., 2006).

Die schlechteren Leistungen der Patienten mit Multipler Sklerose im Vergleich zu den Sprachgesunden in den beiden kognitiv-exekutiv anspruchsvolleren Wechsel-Aufgaben können, ganz allgemein, als Hinweis auf defizitäre Exekutivfunktionen interpretiert werden. Wechsel-Aufgaben erfordern (vgl. auch Abschnitt 2.3) im Vergleich zu simplen Wortgenerierungsaufgaben die Fähigkeit, zwischen verschiedenen Konzepten zu alternieren und sind darüber hinaus mehr von der (zeitlichen) Integration und Organisation von Initiierungs-, Inhibitions und Monitoringprozessen abhängig. Die adäquate Ausführung einer Wechsel-Aufgabe erfordert daher im Vergleich zu den simplen Wortgenerierungsaufgaben mehr exekutive Kontrolle (Henry & Phillips, 2006; Keil & Kaszniak, 2002). Die schlechteren Leistungen der Multiple Sklerose Patienten sind demnach wahrscheinlich auf ein Defizit in den Exekutivfunktionen zurückzuführen. Dies insbesondere, da die Leistungen sowohl in lexikalischen als auch in semantischen Wechsel-Aufgaben beeinträchtigt waren. Die geringere *Anzahl korrekter Nennungen* für die Gruppe der Multiple Sklerose Patienten im Vergleich zu den Sprachgesunden in der

simplen semantischen Aufgabe legt ein zusätzliches semantisches Defizit nahe. Die im Folgenden beschriebenen Unterschiede zwischen den beiden Verlaufsformen der Multiplen Sklerose bestärken diese Vermutung.

Eine zusätzliche Dissoziation der beiden Verlaufsformen der Multiplen Sklerose (MS_SR und MS_CP) war mit den beiden semantischen Wortgenerierungsaufgaben möglich. Sowohl bei der simplen als auch bei der komplexen semantischen Wortgenerierungsaufgabe produzierten die Patienten mit chronisch-progredientem Verlauf signifikant weniger korrekte Nennungen und zeigten damit schlechtere Wortgenerierungsleistungen. In der lexikalischen Wechsel-Aufgabe fanden sich diese Unterschiede zwischen den beiden Patientengruppen nicht. Zusammengenommen können diese Ergebnisse als Hinweis darauf gesehen werden, dass die Unterschiede in den Leistungen zwischen den beiden Patientengruppen eher auf ein semantisches Defizit zurückzuführen sind und sich nicht durch komplexere Anforderungen an exekutive Prozesse erklären lassen. Ähnliche Ergebnisse zeigen auch andere Studien. So fanden beispielsweise Kraus und Kollegen (2005), dass sich verschiedene Verlaufsformen einer Multiplen Sklerose unter anderem anhand von Wortgenerierungsleistungen in semantischen, nicht aber durch Leistungen in lexikalischen Aufgaben dissoziieren lassen. Dies betraf vor allem die Abgrenzung von Patienten mit primär chronisch-progredienter von denen mit sekundär chronisch-progredienter Multipler Sklerose. Patienten mit schubförmigem Verlauf konnten mit Leistungen im verbalen Gedächtnis und anderen Maßen exekutiver Funktionen von denen mit chronisch-progredientem Verlauf der Erkrankung abgegrenzt werden.

Anzumerken ist an dieser Stelle, dass sich (vgl. Tabelle 3.1) die beiden Patientengruppen²⁵ auch bezüglich der kognitiven Leistungen im DemTect (Kalbe et al., 2004) signifikant voneinander unterschieden, mit geringfügig schlechteren kognitiven Leistungen für die Patienten mit chronisch-progredientem Verlauf. Es bleibt daher in einer größeren Stichprobe zu untersuchen, ob sich der Eindruck einer Progredienz in Richtung eines semantischen Defizits bestätigt oder vielmehr generelle kognitive Einschränkungen als Erklärung herangezogen werden müssen.

Darüber hinaus soll an dieser Stelle auch kurz auf die Bedeutung der sprechmotorischen Verlangsamung für die Wortgenerierungsleistungen eingegangen werden. Eine Verlangsamung des Sprechtempos, welches sich auf die Quantität bei der Wortgenerierung auswirken kann (Chiaravalloti & Luca, 2008), konnte bei mehr als zwei Drittel der Patienten mit MS_CP gefunden werden. Demgegenüber war nur bei einem der zehn Patienten mit MS_SR ein leicht verlangsamtes Sprechtempo zu diagnostizieren. Ein negativer Einfluss des Sprechtempos auf die Wortproduktion müsste sich glei-

²⁵DemTect-Werte lagen nur für 15 von 16 der Multiple Sklerose Patienten vor, vgl. Tabelle 3.1.

chermaßen in den verschiedenen Wortgenerierungsaufgaben manifestieren. Dies war in der vorliegenden Studie nicht der Fall. Die schlechteren Wortgenerierungsleistungen der Patienten mit MS_CP im Vergleich zu Patienten MS_SR sind daher nicht auf das reduzierte Sprechtempo zurückzuführen. Der Einfluss sprechmotorischer Defizite auf die Wortgenerierungsleistung kann daher für diese Studie als unwesentlich bezeichnet werden, sollte jedoch bei einer Replikation der Studie mit einer größeren Stichprobe weiter berücksichtigt werden.

Es lässt sich festhalten, dass mit der vorliegenden Studie Hinweise auf quantitative Leistungsunterschiede zwischen Patienten mit MS_SR und Patienten mit MS_CP gefunden werden konnten, welche sich vor allem im Bereich der Semantik zu zeigen schienen. Auch Ergebnisse anderer Untersuchungen weisen auf die Prominenz semantischer Defizite bei Multipler Sklerose hin (Drake et al., 2006). Neben Defiziten bei der Verarbeitung und Enkodierung semantischer Informationen werden insbesondere eine mangelnde Flexibilität des semantischen Systems sowie eine defizitäre Organisation semantischer Inhalte diskutiert. Zudem werden Einschränkungen im semantischen Gehalt beschrieben (Drake et al., 2006; Kraus et al., 2005). Eine Analyse und Interpretation weiterer qualitativer Aspekte der Wortgenerierungsleistung könnte zu einem besseren Verständnis beitragen. Auch die Evaluation der Spontansprache von Patienten mit Multipler Sklerose im Vergleich zu Sprachgesunden kann weitergehende Implikationen anregen. Erste Ergebnisse einer Analyse von Charakteristika der Spontansprache bei Patienten mit MS_SR im Vergleich zu Sprachgesunden zeigten Unterschiede in der Wortfindung, der lexikalischen Dichte sowie in der syntaktischen Komplexität (Böwering, 2012). Weitergehende Analysen, die sich mit dem Zusammenhang von Leistungen in der Spontansprache und bei der Wortgenerierung befassen und insbesondere auch eine Analyse der Formelhaftigkeit sprachlicher Äußerungen einbeziehen, sind in Vorbereitung (Thiele et al., in Vorbereitung). Erste Ergebnisse deuten darauf hin, dass die Sprache von Patienten mit Multipler Sklerose im Vergleich zu Sprachgesunden einen größeren Anteil formelhafter Elemente aufweisen könnte (Thiele et al., in Vorbereitung). Sollte sich die Vermutung eines kausalen Zusammenhangs zwischen reduzierter Wortgenerierungskompetenz und geringerer spontansprachlicher Variabilität bestätigen, könnten auffällige Leistungen in Wortgenerierungsaufgaben ein Indikator für eine detaillierte Analyse der Spontansprache sein. Da der Zeitaufwand für Letztere den der Auswertung von Wortgenerierungsaufgaben deutlich übersteigt, könnte in einigen Fällen gegebenenfalls sogar auf eine Spontanspracheanalyse verzichtet werden.

Darüber hinaus wurden die Wortgenerierungsleistungen in einer Aufgabe zur Verbgenerierung überprüft. Verbgenerierungsleistungen sollen insbesondere sensitiv für kortiko-striatale Pathologien sein und werden beispielswei-

se auch zur Dissoziation beginnender demenzieller Erkrankungen eingesetzt (Östberg et al., 2005; Piatt et al., 1999a,b). Die Ergebnisse der Analysen zeigten, dass sich Patienten mit MS_SR und MS_CP bezüglich ihrer Leistungen in der Wortgenerierung voneinander unterschieden, während sich die Leistungen von Sprachgesunden und Patienten mit MS_SR nicht voneinander abgrenzen ließen. Die Evaluation von Verbgenerierungsleistungen erscheint demnach als geeignetes Maß, um zwischen den verschiedenen Verlaufsformen einer Multiplen Sklerose zu dissoziieren.

Die Produktion von Verben wird typischerweise mit Frontalhirnfunktionen in Verbindung gebracht (Östberg et al., 2007; Piatt et al., 1999b). Es ist jedoch anzunehmen, dass das Produzieren von Verben in einer Wortgenerierungsaufgabe eher semantischen als lexikalischen Abrufstrategien folgt. Hierfür spricht auch, dass Östberg et al. (2007) einen Zusammenhang zwischen reduzierter Verbgenerierungskompetenz und Dysfunktionen im Temporallappen zeigen konnten. Im Gegensatz zu semantischen Wortgenerierungsaufgaben, wie etwa beim Produzieren von Wörtern aus der Kategorie *Tiere*, ist die Kategorie *Verb* jedoch weniger kohärent und erfordert daher einen etwas anderen, strukturierteren Abruf (ähnlich der lexikalischen Wortgenerierung, Östberg et al., 2007). Die Verbgenerierungskompetenz liegt demnach bezüglich der ihr zugrunde liegenden Anforderungen zwischen denen semantischer und lexikalischer Wortgenerierung. Zusammen mit den schlechteren Leistungen der Patienten mit MS_CP in semantischen Wortgenerierungsaufgaben könnte die reduzierte Verbgenerierungsleistung ebenfalls auf ein progredientes semantisches Defizit hinweisen. Die Analyse von *Verb-Clustern* und *Switches* bei der Verbgenerierung könnte weiteren Aufschluss über die Strategien bei der Verbgenerierung liefern und weitergehende Interpretationen ermöglichen.

3.4.2 Interpretation des Anteils korrekter Nennungen sowie der auftretenden Fehler

Darüber hinaus war im Rahmen dieser Studie eine qualitative Evaluation der Wortgenerierungsleistungen von Interesse. Daher wurden der Anteil korrekter Nennungen sowie der Anteil verschiedener Fehlertypen analysiert. Unterschiede im Anteil korrekter Nennungen zeigten sich lediglich bei der lexikalischen Wechsel-Aufgabe. Patienten mit MS_CP produzierten einen signifikant geringeren Anteil korrekter Nennungen im Vergleich zu Sprachgesunden und Patienten mit MS_SR. Die Analyse des Anteils der Fehler ergab nur für die lexikalische und die semantische Wechsel-Aufgabe einen Hinweis auf signifikante Gruppenunterschiede. In der lexikalischen Wechsel-Aufgabe produzierten die Patienten mit MS_CP signifikant mehr einfache Regelbrüche als die Sprachgesunden. In der semantischen Wechsel-Aufgabe zeigten sich signifikante Unterschiede zwischen Sprachgesunden und Patienten mit MS_CP sowie eine Tendenz in Richtung eines Unterschieds zwischen

Patienten mit MS_SR und MS_CP für den Anteil von Kategorieperseverationen.

Kategorieperseverationen können als *stuck in set perseverations* gesehen werden. Sie sind Ausdruck der Schwierigkeit, zwischen verschiedenen semantischen Konzepten zu alternieren. Routinierte semantische Abrufstrategien werden nicht unterbrochen und durch alternative Strategien ersetzt. Das ist ein Verhalten, das typischerweise mit einer Beeinträchtigung des *Supervisory Attentional System* aus dem Modell von Shallice und Kollegen (1996) in Verbindung gebracht wird und somit als Hinweis auf ein kognitiv-exekutives Defizit gesehen werden kann. Das Auftreten von Regelbrüchen und Kategorieperseverationen gilt darüber hinaus als Hinweis auf subkortikale Läsionen und deutet damit auch eher auf eine kognitiv-exekutive Beeinträchtigung hin (Butters et al., 1998).

Auffällig war in allen drei Gruppen außerdem der Anteil von Kategoriefehlern in der semantischen Wechsel-Aufgabe. Bei der Evaluation der Aufgabe zeigte sich, dass insbesondere die Kategorie *Blumen* eine semantisch sehr unscharf definierte Kategorie zu sein scheint. Viele Versuchsteilnehmer, auch Sprachgesunde, generierten bei der Aufgabe nicht nur *Blumen*, sondern fassten die vorgegebene Kategorie weiter und produzierten beispielsweise Pflanzenbestandteile (z.B. *Blatt, Knospe*) oder Ähnliches. Probanden aller Gruppen nannten auffällig häufig sogar verschiedene Baumarten oder Sträucher in dieser Kategorie. So produzierte ein Normsprecher unter anderem die folgenden Nennungen für die Kategorie *Blumen* in der semantischen Wechsel-Aufgabe: *Knospe, Blüte, Kaffeebaum, Bananenbaum etc.* Um eine sichere Identifizierung der Fehlertypen und Interpretation der Fehlerqualität zu gewährleisten, sollte für diese Aufgabe und insbesondere für die Kategorie *Blumen* noch einmal überdacht werden, ob die Instruktionen und Regeln zur Auswertung der Aufgabe gegebenenfalls angepasst werden sollten.

Zusammengefasst lässt sich festhalten, dass die Analyse des Anteils korrekter Nennungen und Fehlertypen bei den im Rahmen dieser Studie untersuchten Stichproben nicht zur Erfassung des zugrunde liegenden Leistungsdefizits geeignet zu sein schien. Unterschiede zwischen den Gruppen zeigten sich hinsichtlich des Anteils korrekter Nennungen nur für die lexikalische Wechsel-Aufgabe, im Hinblick auf den Fehleranteil für beide Wechsel-Aufgaben: bei der lexikalischen Wechsel-Aufgabe waren Unterschiede bei den einfachen Regelbrüchen zu finden; in der semantischen Wechsel-Aufgabe zeigten sich Gruppenunterschiede für den Anteil der Kategorieperseverationen. Zusammengefasst weisen diese Ergebnisse eher auf ein kognitiv-exekutives Defizit hin.

Bei der Betrachtung der Abbildungen 3.6, 3.7, 3.8, 3.9 und 3.10 fällt jedoch auf, dass insbesondere Patienten mit MS_CP vor allem in den beiden Wechsel-Aufgaben numerisch mehr bzw. qualitativ andere Fehler machen.

Eine erneute Analyse der Fehlertypen mit Daten aus einer größeren Stichprobe könnte hier sinnvoll sein. Vor allem bezüglich der Verbgenerierung scheinen weitere qualitative Auswertungen sinnvoll. Bei der Einteilung der Fehlertypen wurde ausschließlich zwischen Perseverationen und Regelbrüchen unterschieden. Die Art der produzierten Regelbrüche variierte jedoch stark. Neben Umschreibungen und Phrasen wurden auch Wörter aus anderen Wortklassen produziert. Daher scheinen weitere Analysen der Regelbrüche auf der Basis qualitativer Unterschiede plausibel.

3.4.3 Fazit

Die Ergebnisse der vorliegenden Studie zeigen, dass sich Patienten mit schubförmig-remittierender und chronisch-progredienter Multipler Sklerose hinsichtlich ihrer Wortgenerierungsleistungen unterscheiden. Insbesondere Leistungen in semantischen Aufgaben sowie bei der Verbgenerierung, gemessen in der *Anzahl korrekter Nennungen*, scheinen gut zwischen den beiden Verlaufsformen einer Multiplen Sklerose zu dissoziieren. Dies kann als Indiz für ein mit Fortschreiten der Erkrankung zunehmendes Defizit in der semantischen Verarbeitung gedeutet werden. Darüber hinaus kann angenommen werden, dass bei Patienten mit Multipler Sklerose unabhängig von der Verlaufsform Defizite in den Exekutivfunktionen bestehen. Dies lässt sich insbesondere anhand der reduzierten Wortgenerierungsleistungen in den beiden Wechsel-Aufgaben erkennen. Das gehäufte Auftreten von Regelbrüchen und Kategorieperseverationen für die Patienten mit chronisch-progredientem Erkrankungsverlauf könnte zudem als Hinweis darauf gewertet werden, dass sich auch die Leistungen in den Exekutivfunktionen bei Progredienz der Erkrankung verschlechtern. Dies entspricht der Annahme von eher umschriebenen (subkortikalen) Defiziten bei schubförmigem Verlauf und sich weitläufig und umfassender manifestierender Defizite bei chronisch-progredienter Multipler Sklerose, entsprechend der Zunahme der Demyelinisierungsherde. Die schlechteren kognitiven Leistungen der Patienten mit chronisch-progredienter Multipler Sklerose im DemTect (Kalbe et al., 2004) können darüber hinaus als Hinweis auf ein globaleres Defizit gedeutet werden.

Einer darüber hinaus gehenden Interpretation der Ergebnisse müssen weitere Studien vorausgehen, die die Wortgenerierungsleistungen von Patienten mit Multipler Sklerose unterschiedlicher Verlaufsformen in größeren Stichproben untersuchen. Weitere Aspekte wie das Manifestationsalter, die Krankheitsdauer, die Schwere der bestehenden neurologischen Beeinträchtigungen, sprechmotorische Beeinträchtigungen sowie kognitive Defizite insgesamt sollten dabei berücksichtigt und in Relation zu den gezeigten Wortgenerierungsleistungen analysiert werden.

Die nun folgende Studie befasst sich mit der Evaluation von Wortgenerierungsleistungen bei Patienten mit Aphasie, Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden. Neben traditionellen Evaluationsmaßen, wie der *Anzahl korrekter Nennungen*, werden weitere zeitliche, sublexikalische sowie lexikalisch-semantische Auswertungsparameter bei der Analyse berücksichtigt. Evaluation und Klassifikation erfolgen auf Basis dieser zusätzlichen sowie der traditionellen Parameter mit *machine learning*-Techniken.

Kapitel 4

Evaluation von Wortgenerierungsleistungen zur Klassifikation von Sprach- und Kommunikationsstörungen mit *machine* *learning*-Techniken

4.1 Einleitung & Fragestellung

Eine verlässliche Dissoziation zwischen Patienten mit Sprach- und/oder Kommunikationsstörungen nach Hirnschädigung (z.B. Aphasie vs. kommunikativ-kognitive Defizite) und Sprachgesunden ist häufig nicht einfach, sondern kann vielmehr als Herausforderung gesehen werden. Insbesondere dann, wenn es um die Diagnose vergleichsweise subtiler Defizite geht oder keine allgemein akzeptierten Standards für eine reliable Diagnosestellung existieren (Lezak et al., 2004).

Während die Diagnosestellung im Fall einer Aphasie relativ zuverlässig auf der Basis etablierter Prozeduren und standardisierter Verfahren möglich ist (vgl. Abschnitt 2.5), fehlen insbesondere zur Diagnose kommunikativ-kognitiver Defizite noch allgemein anerkannte, standardisierte und normierte Verfahren (Duff & Proctor, 2002; Turkstra et al., 2005a,b; Vogel et al., 2010). Darüber hinaus ist auch die Abgrenzung verschiedener Sprach- und Kommunikationsstörungen untereinander, das heißt beispielsweise eine reliable Dis-

soziation zwischen sprachsystematischen und kommunikativ-kognitiven Beeinträchtigungen oder aber die Abgrenzung zwischen entsprechenden Defiziten unterschiedlicher Ätiologien schwierig (Huber et al., 2006; Tesak, 2006; Wehmeyer & Grötzbach, 2006). Eine korrekte Diagnosestellung ist jedoch für die Planung des weiteren Rehabilitationsprozesses wichtig und hängt stark von den identifizierten Charakteristika der diagnostizierten Sprach- und/oder Kommunikationsbeeinträchtigungen ab (vgl. American Speech-Language-Hearing Association, 2005; Jaecks et al., 2012; Parrish, Roth, Roberts & Davie, 2009; Wong et al., 2010). Das Erstellen von Prozeduren, die es ermöglichen, reliabel zwischen pathologischen und normalen Leistungen zu differenzieren, stellt daher eine wichtige Aufgabe für die klinische Praxis dar.

Eine häufig angewendete Methode, insbesondere zur Identifizierung relativ subtiler Beeinträchtigungen, ist die Evaluation von Wortgenerierungsleistungen (siehe hierzu auch Abschnitt 2.3 und Abschnitt 2.5.1). Doch obwohl die Anwendung dieser Methode zur Diagnose sprachlicher und kognitiver Defizite in der klinischen Praxis sehr etabliert ist, ist die Aussagekraft der evaluierten Leistungen häufig begrenzt, was unter anderem auf einen mangelnden Konsens bezüglich relevanter sowie essenzieller Auswertungsparameter zurückzuführen ist.

Traditionell werden in der klinischen Praxis vergleichsweise simple Maße wie die *Anzahl korrekter Nennungen* zur Evaluation von Wortgenerierungsleistungen herangezogen, welche dann im Vergleich zu einer hinsichtlich Alter und/oder Bildung vergleichbaren Referenzstichprobe interpretiert werden (Aschenbrenner et al., 2000). Eine reduzierte *Anzahl korrekter Nennungen* entspricht einer im Vergleich schlechteren Wortgenerierungsleistung. Worauf diese schlechtere Leistung zurückzuführen ist (z.B. auf eine größere Anzahl falscher Nennungen durch eine ausgeprägte Perseverationstendenz und/oder das Auftreten von Regelbrüchen oder auf eine insgesamt reduzierte Anzahl von Nennungen) gibt dieses Maß nicht wieder (Drechsler, 2007).

Über diese traditionelle Evaluation hinausgehende Analysen fokussieren oftmals die Häufigkeit und Art auftretender Fehler (Crowe, 1992; Pekkala, 2012; Raboutet et al., 2010) oder die Anzahl von Nennungen innerhalb (*Clustergröße*) sowie die Anzahl von Wechseln (*Switches*) zwischen Subkategorien (Beatty et al., 1997; Kavé et al., 2011; Troyer et al., 1997). Die Berücksichtigung weiterer sublexikalischer, zeitlicher oder lexikalisch-semantischer Auswertungsparameter spielt bislang bei der Evaluation von Wortgenerierungsleistungen eher eine untergeordnete Rolle. Marklund, Sikstrom, Baath & Nilsson (2009) wendeten ein spezifisches semantisches Maß, die *semantische Kohärenz*, auf Wortgenerierungsdaten an. Die Autoren untersuchten die semantische Kohärenz in zwei Wortgenerierungsaufgaben unterschiedlicher Komplexität bei Personen verschiedenen Alters. Sie verwendeten die *Latent Semantic Analyses* (LSA, Deerwester, Dumais, Furnas, Landauer & Harsh-

man, 1990) um semantische Distanzen zwischen konsekutiven Nennungen zu berechnen und konnten auf Basis dieser Berechnungen altersabhängige Veränderungen in den semantischen Distanzen finden. Die semantische Kohärenz soll auch in der vorliegenden Studie als Auswertungsparameter in die Evaluation der Wortgenerierungsleistungen einbezogen werden.

Als eine mögliche Methode zur Dissoziation zwischen verschiedenen klinischen Populationen und Sprachgesunden wird vermehrt der Einsatz von *machine learning*-Techniken beschrieben (siehe auch Abschnitt 2.6.1 und vgl. z.B. Järvelin, 2008). So verwendete beispielsweise Järvelin (2008) neuronale Netze, um anhand von Daten aus dem Bildbenennen zwischen Patienten mit Aphasie und Sprachgesunden zu dissoziieren. Über 90% der Aphasiepatienten konnten dadurch korrekt klassifiziert werden. Der Autor wies jedoch auch darauf hin, dass es sinnvoll sei, diese Ergebnisse durch Untersuchungen mit anderen Klassifikatoren und Datensets zu erweitern. Während Järvelin (2008) die Klassifikation auf Basis künstlich erzeugter Daten durchführte, werden in der vorliegende Studie reale Patientendaten für die Klassifikation verwendet.

Die folgende Studie befasst sich mit der Evaluation von Wortgenerierungsleistungen bei Patienten mit Aphasie, Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden. Ziel dabei ist, durch Anwendung von *machine learning*-Techniken Variablen zu identifizieren, die eine Dissoziation zwischen diesen Populationen faszilitieren und damit die Erstellung einer reliablen Diagnose unterstützen. Es wird untersucht, ob zusätzliche sublexikalische, lexikalisch-semantische oder zeitliche Variablen, die aus den Wortgenerierungsdaten extrahiert werden, genutzt werden können, um Sprachgesunde von Aphasikern oder Patienten mit kommunikativ-kognitiven Defiziten zu dissoziieren. *machine learning*-Techniken liefern hierbei ein Modell, welches das Risiko einer Fehlklassifikation von Patienten minimieren und als Methode dazu beitragen kann, den Einfluss zusätzlicher Variablen auf die Effizienz von Wortgenerierungsaufgaben zu untersuchen. Ziele sind dabei, (i) eine bessere Klassifikation der beiden Patientengruppen sowie der Sprachgesunden und (ii) ein Vergleich des Einflusses unterschiedlicher Variablen auf den Erfolg der Klassifikation. Dazu wird im Folgenden das methodische Vorgehen näher erläutert, bevor anschließend die Ergebnisse beschrieben und für jede der drei Klassifikationsaufgaben (d.h. *Patienten mit Aphasie vs. Sprachgesunde*, *Patienten mit kommunikativ-kognitiven Defiziten vs. Sprachgesunde* und *Patienten mit Aphasie vs. Patienten mit kommunikativ-kognitiven Defiziten*) diskutiert werden. Abschließend findet sich ein allgemeines Fazit, welches auch auf zukünftige Studien abhebt. Auszüge der folgenden Studie sind veröffentlicht in Gaspers et al. (2012).

4.2 Methode

Zunächst wurden Wortgenerierungsleistungen von Patienten mit Aphasie, Patienten mit kommunikativ-kognitiven Defiziten (*Patienten mit KKD*) und einer Kontrollgruppe sprachgesunder Probanden in insgesamt fünf Wortgenerierungsaufgaben erhoben. Für jede Aufgabe ergab sich eine Liste mit Antworten des jeweiligen Versuchsteilnehmers. Die Antworten der Versuchsteilnehmer wurden transkribiert und entlang einer Zeitachse annotiert. Die daraus resultierende Wortliste bildete die Basis zur Gewinnung von Variablenvektoren, bestehend aus 44 verschiedenen Variablen. Mit unterschiedlichen Strategien zur Variablenauswahl wurden aus diesen 44 Variablen die relevantesten ausgewählt. Dieses reduzierte Variablenset wurde dann verwendet, um verschiedene Klassifikatoren zu trainieren und zu evaluieren. Das genannte Vorgehen wurde bei drei verschiedenen Klassifikationsaufgaben angewendet. Abhängig von der jeweiligen Klassifikationsaufgabe bestand das Ergebnis in der Dissoziation zwischen Patienten mit Aphasie und Sprachgesunden, Patienten mit KKD und Sprachgesunden sowie zwischen Patienten mit KKD und Patienten mit Aphasie. Abbildung 4.1 veranschaulicht den Ablauf von Beginn der Datenerhebung bis hin zur Entscheidung des Klassifikators.



Abbildung 4.1: Workflow: Datenerhebung (*data acquisition*), Annotation und Aufbereitung der Daten (*Annotation/Preprocessing*), Extraktion relevanter Variablen (*Feature extraction*), Variablenauswahl (*Feature selection*), Training, Klassifikator (*Classifier*), Entscheidung bzw. Ergebnis der Klassifikation (*Patient?*); aus Gaspers et al. (2012, S. 211)

4.2.1 Versuchsteilnehmer

Es werden Daten von insgesamt 75 Versuchsteilnehmern aus drei Gruppen präsentiert.²⁶ Alle Versuchsteilnehmer waren monolinguale, deutsche Muttersprachler mit normalem oder korrigiertem Seh- und Hörvermögen. Die Datenerhebung erfolgte in Übereinstimmung mit den 1964 in der Helsinki Deklaration festgelegten ethischen Standards. Alle Teilnehmer gaben nach erfolgter Aufklärung ihre Einwilligung zur Teilnahme an der Studie.

²⁶Bei der Datenerhebung wurde nicht präselektiert, das heißt, die drei Gruppen wurden nicht hinsichtlich Alter, Bildungsgrad und Krankheitsdauer (für die beiden Patientengruppen) gematched, da eine möglichst realitätsnahe Verteilung angestrebt wurde.

Patienten mit Aphasie. Die erste Gruppe bestand aus 20 Patienten (13 Männer, 7 Frauen) mit leichter bis mittelschwerer Aphasie, diagnostiziert mit dem Aachener Aphasie Test (AAT, Huber et al., 1983), im Alter von 40 bis 85 Jahren ($M = 63.65$, $SD = 13.18$). Neurologische Befunde der Ätiologie umfassten neben ischämischen ($n = 14$) und hämorrhagischen ($n = 3$) Insulten auch Schädelhirntraumata ($n = 3$), mit einer Zeit post-onset von 0 bis 280 Monate ($M = 42.95$, $SD = 74.26$). Der Bildungsgrad betrug im Mittel 9.95 Schuljahre ($SD = 1.93$, $Min: 8$, $Max: 14$).

Patienten mit KKD. Die zweite Gruppe bestand aus 20 Patienten (17 Männer, 3 Frauen) mit kommunikativ-kognitiven Defiziten, im Alter von 18 bis 58 Jahren ($M = 41.50$, $SD = 10.98$). Neurologische Befunde der Ätiologie umfassten Schädelhirntraumata ($n = 15$), zerebrale Tumoren ($n = 2$), ischämischen ($n = 1$) und hämorrhagischen ($n = 2$) Insult. Die Krankheitsdauer reichte von 0 bis 175 Monate ($M = 47.50$, $SD = 65.61$). Die Diagnose einer kommunikativ-kognitiven Störung erfolgte auf Basis manifester klinischer Symptome wie z.B. befremdliche Ausdrucksweise, semantische Vagheit und/oder umständliche Formulierungen, stark ausschweifende und/oder sozial nicht akzeptable Äußerungen etc. (vgl. Lê et al., 2011; Martin & McDonald, 2003; Prigatano, 1986).²⁷ Die Diagnose eines KKD wurde gestellt, wenn klinisch Symptome einer kommunikativ-kognitiven Beeinträchtigung von mindestens zwei erfahrenen Beurteilern (z.B. Sprachtherapeut, Arzt) unabhängig voneinander beschrieben wurden. Um auszuschließen, dass eine sprachsystematische Störung Ursache der beobachteten Symptome war, wurden alle Versuchsteilnehmer dieser Gruppe zusätzlich mit dem Aachener Aphasie Screening (*Token Test* und Untertest *Schriftsprache* aus dem AAT, Huber et al., 1983) vorab als nicht-aphasisch eingestuft. Der Bildungsgrad dieser Gruppe betrug im Mittel 10.65 Schuljahre ($SD = 1.50$, $Min = 9$, $Max = 14$).

Sprachgesunde. Die dritte Gruppe bestand aus 35 neurologisch gesunden Kontrollprobanden (17 Männer, 18 Frauen) im Alter von 18 bis 81 Jahren ($M = 51.66$, $SD = 15.86$) und mit einem mittleren Bildungsgrad von 11.83 Schuljahren ($SD = 1.69$, $Min: 8$, $Max: 14$). Anamnestisch bestand bei keinem Versuchsteilnehmer aus dieser Gruppe ein Hinweis auf das Vorliegen einer neurologischen oder psychiatrischen Erkrankung.

²⁷Diese Herangehensweise wurde gewählt da bislang im deutschsprachigen Raum kein allgemein akzeptiertes, standardisiertes und normiertes Verfahren zur Diagnose kommunikativ-kognitiver Defizite existiert.

4.2.2 Material & Durchführung

Die Wortgenerierungsleistungen wurden in insgesamt fünf Aufgaben erhoben. Mit jedem Versuchsteilnehmer wurden zwei simple lexikalische Aufgaben, eine simple semantische Aufgabe sowie zwei komplexere Aufgaben, die einen Wechsel zwischen zwei lexikalischen bzw. zwei semantischen Kategorien erforderten, durchgeführt. Die Aufgabe bestand bei allen Aufgaben für die Versuchsteilnehmer darin, unter Beachtung der vorgegebenen Regeln, innerhalb einer Zeit von 60 Sekunden so viele Wörter wie möglich zu produzieren. Alle Aufgaben wurden stets in derselben Reihenfolge präsentiert. Die Anweisung erfolgte entsprechend der Vorgaben für lexikalische, semantische und Wechsel-Aufgaben aus dem Regensburger Wortflüssigkeitstest (Aschenbrenner et al., 2000).

Die erste lexikalische Aufgabe erforderte das Generieren von Wörtern mit dem Anfangsbuchstaben *A* (z.B. *Apfel, arm, atmen, Affe etc.* – LEX_A), die Zweite das Produzieren von Wörtern mit dem Anfangsbuchstaben *M* (z.B. *Milch, mager, messen, Miete etc.* – LEX_M). In der semantischen Wortgenerierungsaufgabe sollten die Versuchsteilnehmer Wörter der Kategorie *Tiere* (z.B. *Hund, Katze, Pferd, Tiger etc.* – SEM_Tiere) generieren. Die lexikalischen Wechsel-Aufgabe erforderte die alternierende Produktion von Wörtern, die mit den Buchstaben *H* und *T* beginnen (z.B. *Hund, Tomate, hässlich, Tante etc.* – LEX_H/T). In der semantischen Wechsel-Aufgabe sollten alternierend Wörter aus den Kategorien *Kleidung* und *Blumen* (z.B. *Hose, Rose, Kleid, Nelke etc.* – SEM_Kleidung/Blumen) genannt werden. Alle Aufgaben bis auf LEX_A sind Bestandteil des Regensburger Wortflüssigkeitstests (RWT, Aschenbrenner et al., 2000, siehe auch Abschnitt 2.5.1). Die zusätzliche Analyse der Aufgabe LEX_A war durch die Studie von Marklund et al. (2009) motiviert.

Auswertung und Extraktion relevanter Variablen

Ziel der Studie war, neben einer korrekten Dissoziation zwischen den drei zuvor beschriebenen Gruppen, auch die Evaluation des Einflusses spezifischer Variablen. Nach erfolgter Datenerhebung wurden die Nennungen der Versuchsteilnehmer in den einzelnen Aufgaben zunächst annotiert; es resultierte für jeden Versuchsteilnehmer und jede Wortgenerierungsaufgabe eine spezifische Wortliste. Auf Basis bisheriger Befunde aus der Literatur wurden Variablen ausgewählt, die als vielversprechende Kandidaten für die Diskrimination angenommen werden konnten und aus den Wortlisten extrahiert (*feature extraction*). Ausgehend von diesen Variablen wurden die diskriminativsten berechnet (*feature selection*) und anschließend für das Training und die Evaluation verschiedener Klassifikatoren genutzt. Für diese Zwecke wurde die *Weka Data Mining Software* (Version 3.6.2) verwendet (Hall, Frank,

Holmes, Pfahringer, Reutemann, & Witten, 2009).

Aus jeder Wortgenerierungsaufgabe bzw. -leistung resultierte eine Liste L von Nennungen, produziert von einem Versuchsteilnehmer, wobei einige der Nennungen inkorrekte Antworten darstellten. Im Folgenden bezeichnet L_a die Liste aller Nennungen, die in einer Wortgenerierungsaufgabe produziert wurden, während $L_c \subseteq L_a$ die Liste bezeichnet, die nur die korrekten Nennungen enthielt. Die Kodierung korrekter Antworten erfolgte entsprechend der Regeln im Regensburger Wortflüssigkeitstest (Aschenbrenner et al., 2000). Beide Listen L_a und L_c wurden für die Gewinnung der Variablen als wichtig erachtet. Im Folgenden wird ein kurzer Überblick über die relevanten Variablen gegeben. Jede Variable wurde aus einer Wortliste L gewonnen. Für jede L wurden alle Variablen extrahiert. Die Trainingsinstanzen für die anschließende Klassifikation wurden erstellt, indem verschiedene Strategien für die Variablenauswahl verwendet wurden. Es folgt zunächst eine detaillierte Beschreibung der verwendeten Variablen, bevor im Anschluß die zur Variablenauswahl verwendeten Strategien vorgestellt werden.

Traditionelle klinische Maße. Das elementare Variablenset umfasste Variablen, die typischerweise bei der klinischen Analyse von Wortgenerierungsleistungen Anwendung finden: Die Anzahl der Nennungen ($nr_resp_L \triangleq |L|$) für L_a und L_c sowie auch das Verhältnis der beiden Werte zueinander ($perc_correct$).

Analyse der Fehlertypen. Bei der klinischen Analyse von Wortgenerierungsdaten werden verschiedene Fehlertypen unterschieden. In dieser Studie wurde der prozentuale Fehleranteil in L_a für zwei verschiedene Fehlertypen berechnet: Perseverationen ($perc_errors_repetition$) und Regelbrüche ($perc_errors_rule_breaking$). Die Auswertung der Fehlertypen erfolgte entsprechend der Regeln in Aschenbrenner et al. (2000).

Zeitliche Maße. Für die Analyse zeitlicher Variation bei der Wortproduktion wurden eine Reihe zeitlicher Maße eingeführt. Bezüglich der Nennungen in L wurde zunächst die Zeitspanne zwischen aufeinanderfolgenden Nennungen berechnet. Diese Zeitspanne war definiert als Zeit zwischen den Startzeitpunkten von zwei konsekutiven Nennungen sowie als Zeit zwischen dem Startzeitpunkt der letzten Nennung in L und dem Ende der Aufgabe (60sec). Die Startzeitpunkte wurden manuell annotiert. Anschließend wurden als zeitliche Variablen die längste Zeitspanne ($longest_break_L$) und die kürzeste Zeitspanne ($shortest_break_L$) sowie die Varianz der Zeitspanne (var_breaks_L) für L_c und L_a berechnet.

Lexikalisch-semantische Maße. Auf dieser Ebene wurden zwei Arten von Maßen verwendet: ein semantisches Maß (in Anlehnung an Mark-

lund et al., 2009) und ein Wortfrequenzmaß. Hinsichtlich der Semantik wurden zunächst die semantischen Distanzen zwischen konsekutiven Nennungen geschätzt. Daraufhin konnten Mittelwerte aus den erhaltenen Werten berechnet werden, welche das Maß der *semantischen Kohärenz* bildeten. Mit der *Explicit Semantic Analysis* (ESA)²⁸ wurden semantische Ähnlichkeiten berechnet (Gabrilovich & Markovitch, 2007). Mit ESA können Wörter als mehrdimensionale Vektoren repräsentiert werden. Diese setzen sich aus Konzepten zusammen, die aus Wikipedia deriviert wurden. Die semantische Ähnlichkeit zwischen zwei Wörtern berechnete sich auf Basis der Distanzen zwischen den Vektoren die sie repräsentieren (Gabrilovich & Markovitch, 2007). Wortpaare, für die mindestens ein Wort nicht als Konzeptvektor ausgedrückt werden konnte (z.B. Pseudowörter), wurden von der Berechnung ausgeschlossen. Die *Explicit Semantic Analysis* gibt Ergebnisse zwischen 0 und 1 aus, wobei 0 und 1 die geringste bzw. größtmögliche semantische Ähnlichkeit anzeigen. Perseverationen wurden umkodiert von 1 auf 0. Dies hatte den Vorteil, den semantischen Gehalt der generierten Wörter besser abbilden zu können. Für L_c und L_a wurde die semantische Kohärenz ($coherence_L$), die Varianz der semantischen Distanzen ($var_semantic_dist_L$), die größte ($highest_dist_semantic_L$) und die kleinste ($lowest_dist_semantic_L$) semantische Distanz zwischen zwei konsekutiven Nennungen berechnet.

Verschiedene Studien haben einen Wortfrequenzeffekt bei der Wortproduktion in verschiedenen Patientengruppen – insbesondere bei Patienten mit Aphasie – beschrieben (Ellis et al., 1983; Nickels & Howard, 1995). Aus diesem Grund wurden zusätzlich Wortfrequenzmaße verwendet. Da kein Korpus verfügbar war, der einen ausreichend großen Anteil der von den Versuchsteilnehmern produzierten Wörter enthielt, wurde auf das Internet zurückgegriffen. Das heißt, es wurde die Anzahl der in Google angezeigten Ergebnisse bzw. Treffer (*Google Hits*) verwendet. Die *Google Hits* wurden mit der deutschen Version von Google (www.google.de) bestimmt und die resultierenden Werte für die weitere Analyse verwendet. Für L_c und L_a wurde die durchschnittliche Anzahl der *Google Hits* (avg_google_L), der höchste Wert ($highest_google_L$), der niedrigste Wert ($lowest_google_L$) sowie die Varianz der Werte (var_google_L) berechnet.

Sublexikalische Maße. Aufgrund des bei verschiedenen Patientengruppen beschriebenen Wortlängeneffektes (vgl. Nickels & Howard, 2004; Pekkala, 2012) wurde angenommen, dass Patienten mit Sprach- und/oder Kommunikationsstörungen generell weniger komplexe Wörter produzieren als Sprachgesunde. Daher wurde die Anzahl der Silben und Phoneme pro Wort analysiert. Zur Ermittlung der Silben- und Phonemanzahl jedes Wor-

²⁸Research-esa Implementierung, verfügbar unter <http://code.google.com/p/research-esa/>

tes in L wurde die WebCelex²⁹ Datenbank verwendet. Da die Versuchsteilnehmer – insbesondere die Patienten mit Aphasie – eine gewisse Anzahl von Pseudowörtern sowie sehr niederfrequente Wörter produzierten, waren nicht alle Wörter in der WebCelex Datenbank enthalten. Für alle Wörter, die nicht enthalten waren, wurden Silben- und Phonemanzahl manuell annotiert. Für L_c und L_a ergaben sich die folgenden sublexikalischen Maße: kleinste Anzahl von Phonemen ($lowest_phon_L$) und Silben ($lowest_syll_L$), größte Anzahl von Phonemen ($highest_phon_L$) und Silben ($highest_syll_L$), durchschnittliche Anzahl von Phonemen (avg_phon_L) und Silben (avg_syll_L) sowie die Varianz der Anzahl von Phonemen (var_phon_L) und Silben (var_syll_L).

Variablenauswahl

Die Leistung eines Klassifikators ist abhängig vom Verhältnis der Anzahl der Variablen zur Anzahl der für das Training zur Verfügung stehenden Instanzen (d.h. Patienten in den jeweiligen Gruppen). Zusätzliche Variablen können die Leistung eines Klassifikators mindern, wenn die Anzahl der zur Verfügung stehenden Instanzen im Vergleich zu der Anzahl von Variablen relativ gering ist (Jain et al., 2000). Bei einer Anzahl von 44 Variablen war es naheliegend, die Anzahl der Variablen, die zur Klassifikation verwendet werden zu reduzieren; das heißt, die am besten diskriminierenden Variablen sollten vorab identifiziert werden. Zur Einschätzung der diskriminativen Stärke individueller Variablen wurden drei Maße aus dem *machine learning* verglichen: *Chi-square* (im Folgenden: χ^2 , Lehmann & Romano 2005), *Information Gain* (Mitchell, 1997) und *ReliefF* (Kononenko, 1994). Da alle drei Maße vergleichbare Klassifikationsergebnisse erzielten, werden im Folgenden nur die Ergebnisse für χ^2 berichtet.

Mit Weka (Hall et al., 2009) wurden zunächst alle extrahierten Variablen entsprechend ihrer Korrelation mit der Zielkategorie (d.h. Aphasie, KKD, Sprachgesund) anhand der ermittelten χ^2 -Werte in eine Rangfolge (*Ranking*) gebracht. Dieses *Ranking* wurde dann als Basis für die Variablenauswahl bei der folgenden Klassifikation genutzt. Für die Auswahl der Variablen (*feature selection*, vgl. nochmal Abbildung 4.1) wurden verschiedene Strategien verwendet. Da die klinische Auswertung von Wortgenerierungsaufgaben traditionell auf Basis der *Anzahl korrekter Nennungen* vorgenommen wird, wurde die Variable $nr_resp_{L_c}$ als Baseline verwendet. Basierend auf dem χ^2 -*Ranking* wurden zusätzliche Variablen in die Analyse einbezogen. Außerdem wurde die Klassifikationsgüte bei Ausschluss der Baseline $nr_resp_{L_c}$ betrachtet. Zusammengefasst wurden die folgenden Strategien bei der Variablenauswahl angewendet:

²⁹<http://celex.mpi.nl/>

- Baseline (BL): die Variable $nr_resp_{L_c}$ (Baseline-Variable) wird als einzige Variable verwendet
- Baseline plus weitere Variablen, die entsprechend ihres χ^2 -Wertes ausgewählt werden (BL+CHI): der Baseline werden Variablen entsprechend der Abnahme ihrer χ^2 -Werte hinzugefügt, bis sich die *accuracy*³⁰ verringert
- Variablenauswahl entsprechend der χ^2 -Werte (CHI): Variablen werden entsprechend der Abnahme ihrer χ^2 -Werte ausgewählt bis sich die *accuracy* verringert
- Variablenauswahl entsprechend der χ^2 -Werte unter Ausschluss der Baseline (CHI-BL): der Unterschied zu CHI besteht darin, dass die Variablen $nr_resp_{L_c}$ und $nr_resp_{L_a}$ nicht berücksichtigt werden

Das Ziel der Variablenauswahl bestand darin, eine korrekte Klassifikation zu erzielen und zusätzliche Variablen nur dann einzubeziehen, wenn sie die Klassifikation bereicherten. Daher wurde die Variablenauswahl auf Basis der *accuracy*, das heißt anhand des Prozentsatzes korrekt klassifizierter Instanzen für BL+CHI, CHI und CHI-BL optimiert.

4.3 Ergebnisse

Mit Weka (Hall et al., 2009) wurden, basierend auf den zuvor ermittelten Variablen, verschiedene Klassifikatoren trainiert und evaluiert. Die im Folgenden gelisteten Klassifikatoren wurden ausgewählt, da sie unterschiedliche Paradigmen modernster *machine learning*-Algorithmen repräsentieren:

- 1-Nearest-Neighbour classifier (1NN; Aha & Kibler, 1991), Nearest-Neighbour-Klassifikator
- Logistic Model Tree (LMT; Landwehr, Hall & Frank, 2005), Entscheidungsbaum
- Multi Layer Perceptron (MLP; Haykin, 1999), Klassifikation mit neuronalen Netzen
- Naive Bayes (NB; John & Langley, 1995), probabilistischer Klassifikator
- Support Vector Machines (SVM, linearer Kernel mit Platt's *minimal optimization algorithm for training*; Platt, 1998), Maximum-Margin-Klassifikator

³⁰ *Accuracy* heißt im Zusammenhang mit *machine learning* so viel wie Klassifikationsgenauigkeit oder Treffgenauigkeit. Da sich im deutschen Sprachraum die Verwendung des englischen Ausdrucks etabliert hat, wird er auch hier vorrangig verwendet.

4.3.1 Klassifikation

Alle Klassifikatoren wurden mit 10-facher Kreuzvalidierung evaluiert (siehe hierzu auch Abschnitt 2.6). Für jeden Durchgang wurde die Variablenauswahl entsprechend der verschiedenen Strategien auf den Trainingsdaten durchgeführt. Die Klassifikatoren wurden ihrerseits unter Verwendung der resultierenden Variablensets trainiert und anschließend auf Basis der Testdaten evaluiert (die Datensätze wurden gefiltert, damit sie nur die gewählten Variablen enthalten). Anschließend wurde die *accuracy* der einzelnen Klassifikatoren verglichen.

Für die Stichprobe dieser Studie gab es drei verschiedene Klassifikationsaufgaben: (i) Patienten mit Aphasie versus Sprachgesunde, (ii) Patienten mit KKD versus Sprachgesunde und (iii) Patienten mit Aphasie versus Patienten mit KKD. Im Folgenden werden die Ergebnisse für jede der drei Klassifikationsaufgaben dargestellt. In den Tabellen sind die jeweils besten Ergebnisse (= höchste *accuracy*) für die Klassifikatoren bzw. Aufgaben hervorgehoben.

Klassifikation: Patienten mit Aphasie vs. Sprachgesunde

Die Ergebnisse der Klassifikation von Patienten mit Aphasie gegenüber den Sprachgesunden sind in Tabelle 4.1 abgebildet. Die Tabelle gibt die *accuracy* gemittelt über die fünf Wortgenerierungsaufgaben an.

Tabelle 4.1: *Accuracy gemittelt über alle Aufgaben (Patienten mit Aphasie vs. Sprachgesunde)*

Klassifikator	BL	BL+CHI	CHI	CHI-BL
1NN	87.21	88.32	88.32	85.03
LMT	92.69	91.59	90.5	85.01
MLP	90.13	89.41	88.68	85.4
NB	92.69	90.5	90.13	85.05
SVM	92.69	92.69	93.06	86.49

Von den insgesamt fünf Klassifikatoren zeigten drei (LMT, NB und MLP) die besten Ergebnisse für die Baseline-Variable $nr_resp_{L_c}$ (BL); zusätzliche Variablen verringerten die *accuracy* geringfügig. Das beste Ergebnis (mit einer *accuracy* von 93.06%) wurde für den Klassifikator SVM mit der Strategie CHI erzielt, einer Strategie, welche die Variablen anhand des χ^2 -Rankings auswählt. Dieses stellt eine geringfügige Verbesserung im Vergleich zur Baseline in Höhe von 0.37% dar. Der SVM Klassifikator erzielte generell die besten Ergebnisse. Demgegenüber wurden, mit einer *accuracy* in Höhe von bis zu 88.32%, die schwächsten Ergebnisse für den 1NN Klassifikator beobachtet. Alle Klassifikatoren erzielten auch dann noch sehr gute Ergebnisse (\approx 85-86%) bei der Klassifikation von Patienten mit Aphasie

und Sprachgesunden, wenn die Variablen $nr_resp_{L_c}$ und $nr_resp_{L_a}$ ausgeschlossen waren, das heißt mit der Strategie BL-CHI.

Tabelle 4.2: *Accuracy gemittelt über alle Klassifikatoren (Patienten mit Aphasie vs. Sprachgesunde)*

Aufgabe	BL	BL+CHI	CHI	CHI-BL
LEX_A	88.52	88.15	88.15	84.07
LEX_M	87.27	88.36	87.64	84.36
SEM_Tiere	91.27	89.82	89.09	84.36
LEX_H/T	97.45	95.64	95.64	89.82
SEM_Kleidung/ Blumen	90.91	90.55	90.18	84.36

Tabelle 4.2 zeigt die *accuracy* für die einzelnen Wortgenerierungsaufgaben gemittelt über alle Klassifikatoren. Für drei von fünf Aufgaben (LEX_A, SEM_Tiere und LEX_H/T) wurden die besten Ergebnisse erzielt, wenn nur die Baseline $nr_resp_{L_c}$ verwendet wurde, während für die verbliebenen zwei Aufgaben (LEX_M und SEM_Kleidung/Blumen) die besten Ergebnisse mit der Strategie BL+CHI erreicht werden konnten. Das beste Ergebnis unter den fünf Aufgaben wurde mit der Aufgabe LEX_H/T erzielt. Obwohl der höchste Wert in der Aufgabe LEX_H/T bei Verwendung der Baseline gefunden wurde (97.45%), erzielte auch die Strategie *CHI-BL* ein Klassifikationsergebnis, das immer noch höher war, als ein mit den simplen lexikalischen Aufgaben erzielt. Im Folgenden werden die Unterschiede in der *accuracy* für verschiedene Aufgaben-Typen, das heißt für lexikalische (LEX_A und LEX_M), semantische (SEM_Tiere) und Wechsel-Aufgaben (LEX_H/T und SEM_Kleidung/Blumen) separat beschrieben.

Tabelle 4.3: *Accuracy gemittelt über die simplen lexikalischen Aufgaben (Patienten mit Aphasie vs. Sprachgesunde)*

Klassifikator	BL	BL+CHI	CHI	CHI-BL
1NN	82.57	88.07	88.07	83.49
LMT	89.92	88.08	87.17	80.71
MLP	87.16	87.16	87.16	87.14
NB	89.92	87.16	86.25	85.35
SVM	89.92	90.82	90.82	84.42

Tabelle 4.3 zeigt die *accuracy* für alle Klassifikatoren gemittelt über die beiden simplen lexikalischen Wortgenerierungsaufgaben LEX_A und LEX_M. Während die *accuracy* für die Klassifikatoren SVM und 1NN durch zusätzliche Variablen verbessert wurde, verringerte sich die Klassifikations-

4 Evaluation von Wortgenerierungsleistungen zur Klassifikation von Sprach- und Kommunikationsstörungen mit *machine learning*-Techniken

genauigkeit der Klassifikatoren LMT und NB hierdurch. Für den Klassifikator MLP blieb die *accuracy*, unabhängig von der bei der Variablenauswahl verwendeten Strategie, konstant.

Tabelle 4.4: *Accuracy in der simplen semantischen Aufgabe (Patienten mit Aphasie vs. Sprachgesunde)*

Klassifikator	BL	BL+CHI	CHI	CHI-BL
1NN	89.09	83.64	83.64	80.0
LMT	92.73	96.36	94.55	90.91
MLP	89.09	87.27	83.64	83.64
NB	92.73	90.91	90.91	81.82
SVM	92.73	90.91	92.73	85.45

Tabelle 4.5: *Accuracy gemittelt über die Wechsel-Aufgaben (Patienten mit Aphasie vs. Sprachgesunde)*

Klassifikator	BL	BL+CHI	CHI	CHI-BL
1NN	90.91	90.91	90.91	89.09
LMT	95.46	92.73	91.82	86.36
MLP	93.64	92.73	92.73	84.54
NB	95.46	93.64	93.64	86.37
SVM	95.46	95.46	95.46	89.09

Die Tabellen 4.4 und 4.5 zeigen die Klassifikationsergebnisse für die semantische Wortgenerierungsaufgabe (SEM_Tiere) bzw. die Wechsel-Aufgaben (LEX_H/T und SEM_Kleidung/Blumen). Für die semantische Aufgabe konnten für vier von fünf Klassifikatoren (1NN, MLP, NB, SVM) die besten Ergebnisse mit der Baseline $nr_resp_{L_c}$ erzielt werden. Bei den Wechsel-Aufgaben erzielte die Baseline-Strategie (BL) für alle Klassifikatoren die besten Ergebnisse.

Insgesamt lässt sich feststellen, dass der Klassifikator SVM für die Dissoziation zwischen Patienten mit Aphasie und Sprachgesunden am besten geeignet zu sein scheint. Bei der Evaluation gemittelt über alle Wortgenerierungsaufgaben (vgl. Tabelle 4.1) konnten die besten Ergebnisse mit der Baseline-Strategie (BL) erzielt werden. Die beste *accuracy* wurde bei der Evaluation der Wechsel-Aufgaben erreicht. Hier zeigte sich somit die alleinige Berücksichtigung der traditionell bei der klinischen Analyse verwendeten Variable $nr_resp_{L_c}$ als geeignet: gemittelt über beide Aufgaben wurde eine *accuracy* in Höhe von bis zu 95.46% gefunden.

Klassifikation: Patienten mit KKD vs. Sprachgesunde

Die Ergebnisse der Klassifikation von Patienten mit KKD und Sprachgesunden, gemittelt über alle Wortgenerierungsaufgaben, sind in Tabelle 4.6 aufgeführt. Die *accuracy* ist für alle Klassifikatoren angegeben. Die Strategien BL+CHI, CHI und CHI-BL zeigten eine beträchtliche Verbesserung im Vergleich zur Baseline-Strategie für alle Klassifikatoren. Für drei von fünf Klassifikatoren (1NN, MLP und LMT) wurde das beste Ergebnis erzielt, wenn die Baseline $nr_resp_{L_c}$ sowie die Variable $nr_resp_{L_a}$ ausgeschlossen wurden (d.h. CHI-BL). Das beste Ergebnis wurde für den Klassifikator 1NN mit einer *accuracy* in Höhe von 78.55% bei Verwendung der CHI-BL Strategie erzielt. Die Verbesserung gegenüber der Baseline betrug dabei 15.64%. Während der SVM Klassifikator die besten Ergebnisse bei der Klassifikation von Patienten mit Aphasie und Sprachgesunden zeigte, erzielte er die schlechtesten Ergebnisse bei der Klassifikation von Patienten mit KKD und Sprachgesunden. Das entgegengesetzte Bild zeigte sich für den Klassifikator 1NN.

Tabelle 4.6: *Accuracy gemittelt über alle Aufgaben (Patienten mit KKD vs. Sprachgesunde)*

Klassifikator	BL	BL+CHI	CHI	CHI-BL
1NN	62.91	68.73	74.55	78.55
LMT	66.18	73.09	76.36	77.46
MLP	68.73	73.09	74.55	76.73
NB	67.64	73.82	75.64	74.55
SVM	64.0	68.73	67.64	68.0

Tabelle 4.7 zeigt die *accuracy* für die einzelnen Wortgenerierungsaufgaben, gemittelt über alle Klassifikatoren. Die beste *accuracy* wurde mit der Strategie CHI-BL für die Wechsel-Aufgaben erzielt; das beste Ergebnis in Höhe von 78.18% zeigte die Aufgabe SEM_Kleidung/Blumen. Für die semantische Aufgabe SEM_Tiere wurden die besten Ergebnisse mit der Strategie BL+CHI erreicht. In den beiden simplen lexikalischen Aufgaben führten die Strategien CHI und CHI-BL zu den besten und vergleichbar guten Ergebnissen.

Tabelle 4.8 zeigt die Klassifikationsergebnisse gemittelt über die beiden einfachen lexikalischen Aufgaben (LEX_A und LEX_M). Die Strategien CHI und CHI-BL erzielten vergleichbare *accuracy*-Werte und damit gleichzeitig die besten für alle Klassifikatoren. Das beste Ergebnis konnte für den Klassifikator 1NN mit einer *accuracy* in Höhe von 77.28% gefunden werden. Im Gegensatz dazu wurde für die simple semantische Aufgabe (SEM_Tiere; Tabelle 4.9) das beste Ergebnis mit der BL+CHI Strategie

4 Evaluation von Wortgenerierungsleistungen zur Klassifikation von Sprach- und Kommunikationsstörungen mit *machine learning*-Techniken

Tabelle 4.7: *Accuracy gemittelt über alle Klassifikatoren (Patienten mit KKD vs. Sprachgesunde)*

Aufgabe	BL	BL+CHI	CHI	CHI-BL
LEX_A	61.09	70.55	76.36	76.36
LEX_M	60.0	60.36	70.55	70.55
SEM_Tiere	68.0	76.73	71.64	73.09
LEX_H/T	71.27	73.82	74.55	77.09
SEM_Kleidung/ Blumen	69.09	76.0	75.64	78.18

Tabelle 4.8: *Accuracy gemittelt über die simplen lexikalischen Aufgaben (Patienten mit KKD vs. Sprachgesunde)*

Klassifikator	BL	BL+CHI	CHI	CHI-BL
1NN	57.28	57.27	77.28	77.28
LMT	60.0	67.27	76.37	76.37
MLP	61.82	68.18	75.46	75.46
NB	60.0	68.19	71.82	71.82
SVM	63.64	66.37	66.37	66.37

und dem MLP Klassifikator ($accuracy = 83.64\%$) erreicht. Für die beiden Wechsel-Aufgaben (LEX_H/T und SEM_Kleidung/Blumen; Tabelle 4.10) wurden für drei von fünf Klassifikatoren (1NN, LMT und MLP) die besten Ergebnisse mit der Strategie CHI-BL erzielt. Den höchsten Wert ($accuracy = 81.82\%$) erzielte der Klassifikator 1NN.

Tabelle 4.9: *Accuracy in der simplen semantischen Aufgabe (Patienten mit KKD vs. Sprachgesunde)*

Klassifikator	BL	BL+CHI	CHI	CHI-BL
1NN	58.18	80.0	74.55	74.55
LMT	72.73	80.0	76.36	80.0
MLP	72.73	83.64	70.91	74.55
NB	72.73	76.36	76.36	74.55
SVM	63.64	63.64	60.0	61.82

Von den getesteten Klassifikatoren scheint 1NN am besten geeignet, um zwischen Patienten mit KKD und Sprachgesunden zu dissoziieren. Bei der Auswertung über alle Wortgenerierungsaufgaben konnten die besten Ergebnisse für die Strategie CHI-BL gefunden werden. Wie auch bei der Klassifikation von Patienten mit Aphasie und Sprachgesunden scheinen bessere Ergebnisse erreicht werden zu können, wenn nur bestimmte Aufgaben evaluiert werden. Die Analyse der Wechsel-Aufgaben scheint auch bei der Dissoziation

Tabelle 4.10: *Accuracy gemittelt über die Wechsel-Aufgaben (Patienten mit KKD vs. Sprachgesunde)*

Klassifikator	BL	BL+CHI	CHI	CHI-BL
1NN	70.91	74.55	71.82	81.82
LMT	69.09	75.46	76.36	77.28
MLP	73.63	72.73	75.46	79.1
NB	72.73	78.18	79.09	77.27
SVM	64.55	73.63	72.72	72.72

zwischen Patienten mit KKD und Sprachgesunden vielversprechend zu sein. Mit den Wechsel-Aufgaben wurde eine *accuracy* in Höhe von bis zu 81.82% erzielt. Aber auch mit der simplen semantischen Aufgabe konnten sehr gute Klassifikationsergebnisse erreicht werden. Anders als bei der Klassifikation von Patienten mit Aphasie und Sprachgesunden scheint bei der Dissoziation zwischen Patienten mit KKD und Sprachgesunden die Strategie am besten geeignet zu sein, die andere als die traditionell verwendeten Variablen berücksichtigt (CHI-BL). Die Ergebnisse deuten darauf hin, dass es bei der Evaluation von Wortgenerierungsleistungen bei Patienten mit KKD sinnvoll sein könnte, Variablen wie $nr_resp_{L_c}$ oder $nr_resp_{L_a}$ nicht zu berücksichtigen und an ihrer Stelle andere Variablen in die Analyse einzubeziehen.

Klassifikation: Patienten mit KKD vs. Patienten mit Aphasie

Tabelle 4.11 zeigt die Klassifikationsergebnisse für die einzelnen Wortgenerierungsaufgaben gemittelt, über alle Klassifikatoren. Wie in den zuvor präsentierten Klassifikationen wurde auch hier die beste *accuracy* für die beiden Wechsel-Aufgaben erzielt, in diesem Fall für die Aufgabe LEX_H/T mit einer *accuracy* in Höhe von 90.5% bei Anwendung der Baseline-Strategie (BL), sowie für die Strategien BL+CHI und CHI.

Tabelle 4.11: *Accuracy gemittelt über alle Klassifikatoren (Patienten mit KKD vs. Patienten mit Aphasie)*

Aufgabe	BL	BL+CHI	CHI	CHI-BL
LEX_A	81.03	78.46	75.9	74.36
LEX_M	79.5	78.0	73.0	74.5
SEM_Tiere	78.0	74.5	75.0	71.5
LEX_H/T	90.5	90.5	90.5	84.0
LEX_Kleidung/ Blumen	82.0	80.0	73.0	72.5

Die Ergebnisse der Klassifikation von Patienten mit KKD und Patienten mit Aphasie, gemittelt über alle Wortgenerierungsaufgaben, sind in Tabelle

4 Evaluation von Wortgenerierungsleistungen zur Klassifikation von Sprach- und Kommunikationsstörungen mit *machine learning*-Techniken

Tabelle 4.12: *Accuracy gemittelt über alle Aufgaben (Patienten mit KKD vs. Patienten mit Aphasie)*

Klassifikator	BL	BL+CHI	CHI	CHI-BL
1NN	76.87	76.37	76.37	76.38
LMT	83.4	80.9	74.35	74.33
MLP	84.44	79.88	77.88	73.36
NB	83.91	82.41	79.9	77.9
SVM	82.41	81.9	78.9	74.88

4.12 abgebildet. Das beste Ergebnis wurde für alle Klassifikatoren mit der Baseline-Strategie (BL) erreicht. Durch das Einbeziehen zusätzlicher Variablen verringerte sich die *accuracy*. Das beste Ergebnis erzielte der MLP Klassifikator mit einer *accuracy* in Höhe von 84.44%.

Von den getesteten Klassifikatoren scheint MLP am besten geeignet, um zwischen Patienten mit KKD und Patienten mit Aphasie zu dissoziieren. Wie bei den zuvor beschriebenen Klassifikationen (Patienten mit Aphasie vs. Sprachgesunde und Patienten mit KKD vs. Sprachgesunde), können auch hier die besten Ergebnisse erzielt werden, wenn die verschiedenen Wortgenerierungsaufgaben individuell evaluiert werden. Für die Dissoziation zwischen Patienten mit KKD und Patienten mit Aphasie scheint die Evaluation der Aufgabe LEX_H/T die beste Wahl zu sein. Wie bei der Klassifikation von Patienten mit Aphasie und Sprachgesunden, scheint auch bei der Unterscheidung zwischen Patienten mit Aphasie und Patienten mit KKD die alleinige Betrachtung der traditionellen Baseline-Variable *Anzahl korrekter Nennungen* für die Dissoziation geeignet zu sein.

4.3.2 Variablenanalyse

In den vorausgehenden Abschnitten konnte gezeigt werden, dass die Auswahl von Variablen anhand ihrer χ^2 -Werte im Vergleich zur Baseline zu besseren Klassifikationsergebnissen bei der Dissoziation zwischen Patienten mit KKD und Sprachgesunden führt. Auch wenn dies nicht für die Klassifikation von Patienten mit Aphasie und Sprachgesunden zutrifft, konnten bei Verwendung des χ^2 -Rankings ebenfalls annehmbare Ergebnisse bei der Dissoziation zwischen Aphasikern und Sprachgesunden erzielt werden.

Es folgt nun eine weitere Analyse der untersuchten Variablen. Dabei wurden die verschiedenen Variablen evaluiert indem ihre Korrelation mit der Ziel-Kategorie (d.h. Patienten mit Aphasie, Patienten mit KKD, Sprachgesunde) anhand des in Abschnitt 4.2.2 eingeführten χ^2 -Maßes verglichen wurde. Mit Weka (Hall et al., 2009) wurden die χ^2 -Werte für alle Variablen in den drei Klassifikationsaufgaben berechnet. Für jede Klassifikation wurde

zunächst der χ^2 -Wert für alle Variablen in jeder Wortgenerierungsaufgabe berechnet. Anschließend wurden die Werte über alle Wortgenerierungsaufgaben gemittelt und die Variablen entsprechend ihrer χ^2 -Werte sortiert (Ranking).

Tabelle 4.13: *Liste der besten zehn Variablen gemittelt über alle Aufgaben (Patienten mit Aphasie vs. Sprachgesunde)*

Variable	χ^2 -Wert
$nr_resp_{L_c}$	39.56
$nr_resp_{L_a}$	38.45
$shortest_break_{L_c}$	37.9
$shortest_break_{L_a}$	32.04
$var_breaks_{L_c}$	30.53
$longest_break_{L_c}$	29.78
$var_breaks_{L_a}$	29.47
$longest_break_{L_a}$	28.81
$var_syll_{L_c}$	23.99
$lowest_google_{L_c}$	22.48

Tabelle 4.13 zeigt eine Liste der besten zehn Variablen für die Klassifikation von Patienten mit Aphasie und Sprachgesunden, sortiert entsprechend ihres χ^2 -Rankings. Die Variable $nr_resp_{L_c}$, die traditionell zur Evaluation von Wortgenerierungsaufgaben verwendet wird, erzielte das beste Ergebnis ($\chi^2 = 39.56$); die Variable, die die Gesamtzahl aller Nennungen einbezieht, $nr_resp_{L_a}$, erzielte nur ein geringfügig schlechteres ($\chi^2 = 38.45$). Die sechs auf zeitlichen Charakteristika basierenden Variablen nahmen die Plätze drei bis acht auf der Liste ein. Damit bildet die Gruppe der zeitbasierten Variablen die mit den meisten Vertretern unter den besten zehn. Die Variable $shortest_break_{L_c}$ ($\chi^2 = 37.9$) erzielte die höchsten Werte. Zusätzlich zu den zeitlichen und den traditionellen Variablen fanden sich noch die Variablen $var_syll_{L_c}$ ($\chi^2 = 23.99$) und $lowest_google_{L_c}$ ($\chi^2 = 22.48$) unter den besten zehn. Für die Variablen x wurde sowohl x_{L_a} als auch x_{L_c} während der Variablenextraktion berechnet. Die Liste zeigt, dass eine Variable x_{L_c} immer einen höheren Wert erzielte als die entsprechende Variable x_{L_a} .

Tabelle 4.14 zeigt eine Liste der besten zehn Variablen für die Klassifikation von Patienten mit Aphasie und Patienten mit KKD, sortiert entsprechend ihres χ^2 -Rankings. Auch hier erzielte die Variable $nr_resp_{L_c}$ ($\chi^2 = 21.74$) die besten Werte, gefolgt von $nr_resp_{L_a}$ ($\chi^2 = 21.0$) mit nur einem geringfügig kleineren χ^2 -Wert. Obwohl die Variablen $nr_resp_{L_c}$ und $nr_resp_{L_a}$ in beiden Klassifikationen (d.h. Patienten mit Aphasie vs. Sprachgesunde und Patienten mit Aphasie vs. Patienten mit KKD) die bes-

4 Evaluation von Wortgenerierungsleistungen zur Klassifikation von Sprach- und Kommunikationsstörungen mit *machine learning*-Techniken

Tabelle 4.14: *Liste der besten zehn Variablen gemittelt über alle Aufgaben (Patienten mit Aphasie vs. Patienten mit KKD)*

Variable	χ^2 -Wert
<i>nr_resp</i> _{L_c}	21.74
<i>nr_resp</i> _{L_a}	21.0
<i>longest_break</i> _{L_c}	19.04
<i>longest_break</i> _{L_a}	18.77
<i>var_syll</i> _{L_c}	16.92
<i>var_phon</i> _{L_a}	16.63
<i>var_breaks</i> _{L_a}	15.63
<i>shortest_break</i> _{L_c}	14.3
<i>var_syll</i> _{L_a}	14.1
<i>highest_syll</i> _{L_c}	12.9

ten Werte erreichten, war der χ^2 -Wert der beiden Variablen bei der Klassifikation von Patienten mit Aphasie und Patienten mit KKD nur etwa halb so hoch wie bei der Klassifikation von Patienten mit Aphasie und Sprachgesunden. Anders als bei der Klassifikation von Patienten mit Aphasie und Sprachgesunden waren bei dieser Klassifikation nur vier der sechs zeitlichen Variablen in der Liste der besten zehn enthalten. Variablen, die auf der längsten Zeitspanne basieren (nicht wie bei der Klassifikation von Patienten mit Aphasie und Sprachgesunden, Variablen der kürzesten Zeitspanne) erzielten die besten χ^2 -Werte: *longest_break*_{L_c}, $\chi^2 = 19.04$ und *longest_break*_{L_a}, $\chi^2 = 18.77$. Zusätzlich zu den zeitlichen und den traditionellen Variablen enthielt die Liste der besten zehn vier sublexikalische Variablen.

Bei der Klassifikation von Patienten mit Aphasie und Patienten mit KKD wurden generell geringere χ^2 -Werte im Vergleich zur Klassifikation von Patienten mit Aphasie und Sprachgesunden erreicht. Sogar die Variable, die bei der Klassifikation von Patienten mit Aphasie und Sprachgesunden an der letzten Stelle gelistet war (*lowest_google*_{L_c}, $\chi^2 = 22.48$), erreichte höhere χ^2 -Werte als die am besten gelistete Variable (*nr_resp*_{L_c}, $\chi^2 = 21.74$) bei der Klassifikation von Patienten mit Aphasie und Sprachgesunden. Während bei der Klassifikation von Patienten mit Aphasie und Sprachgesunden bei einer Variable x der Wert x_{L_c} immer über der entsprechenden Variable x_{L_a} gelistet war, konnte dies für die Klassifikation von Patienten mit Aphasie und Patienten mit KKD nicht gefunden werden. So waren beispielsweise die Variablen *var_phon*_{L_a} und *var_breaks*_{L_a} in der Liste der besten zehn Variablen enthalten, aber die entsprechenden Variablen aus L_c nicht.

Tabelle 4.15 zeigt eine Liste der besten Variablen für die Klassifikation von Patienten mit KKD und Sprachgesunden, sortiert entsprechend ihres χ^2 -Rankings. In der Tabelle sind alle Variablen gelistet, die einen

Tabelle 4.15: *Liste der besten Variablen gemittelt über alle Aufgaben (Patienten mit KKD vs. Sprachgesunde)*

Variable	χ^2 -Wert
<i>shortest_break_{L_c}</i>	19.55
<i>shortest_break_{L_a}</i>	19.53
<i>nr_respl_a</i>	5.24
<i>nr_respl_c</i>	5.06
<i>perc_errors_rule_breaking</i>	2.36
<i>coherence_{L_a}</i>	2.36

χ^2 -Wert größer Null erzielen. Die besten χ^2 -Werte erzielte die Variable *shortest_break_{L_c}* ($\chi^2 = 19.55$), gefolgt von *shortest_break_{L_a}* ($\chi^2 = 19.53$). Traditionell bei der Evaluation von Wortgenerierungsaufgaben verwendete Variablen waren an dritter und vierter Stelle gelistet, erzielten jedoch lediglich χ^2 -Werte in der Höhe von etwa einem Viertel der beiden als beste gelistete Variablen.

4.4 Diskussion

Für klinische Zwecke ist es von großer Wichtigkeit, sowohl unterschiedliche Formen von Sprach- und/oder Kommunikationsstörungen als auch Patienten mit subtilen Beeinträchtigungen von Sprachgesunden reliabel dissoziieren zu können. Eine solche Dissoziation ist wichtig, um die Effizienz von Diagnostik und Therapie zu verbessern und hat darüber hinaus als langfristiges Ziel einen sozialen und ökonomischen Gewinn für das öffentliche Gesundheitssystem.

Die beschriebene Studie untersuchte die Anwendbarkeit von *machine learning*-Techniken auf Wortgenerierungsdaten bei Patienten mit Aphasie, Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden. Das Ziel war dabei, Variablen zu identifizieren, die sowohl eine reliable Dissoziation zwischen Patienten mit Sprach- und Kommunikationsstörungen (Patienten mit Aphasie und Patienten mit kommunikativ-kognitiven Defiziten) und Sprachgesunden, als auch zwischen den beiden Patientengruppen selbst ermöglichen.

Es wurden Daten präsentiert, die einen großen Einfluss verschiedener Variablen auf die Evaluation von Wortgenerierungsdaten nahelegen. Während die Analyse anderer als traditionell in der klinischen Praxis verwendeter Variablen die Klassifikation von Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden verbesserte, konnte dies nicht für die Klassifikation von Patienten mit Aphasie und Sprachgesunden bestätigt werden. Anzumerken ist, dass die relativ geringe Anzahl zur Verfügung stehender

Trainingsdaten Training und Evaluation der Klassifikatoren erschwert haben könnte, wenn nach Anwendung der Auswahlstrategie mehrere Variablen in einem Set enthalten waren. Wie in Abschnitt 4.2.2 bereits erwähnt, können zusätzliche Variablen die Performanz eines Klassifikators verringern, wenn die Anzahl der Instanzen im Vergleich zur Anzahl der Variablen relativ gering ist (Jain et al., 2000). Denkbar ist, dass die Verwendung zusätzlicher Variablen auch die Klassifikation von Patienten mit Aphasie und Sprachgesunden im Vergleich zur Baseline verbessert, wenn ein größeres Set von Trainingsdaten zur Verfügung steht. Eine genauere Parameter-Bewertung der Klassifikatoren ist jedoch erst möglich, wenn eine vergleichbar große Anzahl von Variablen und Trainingsinstanzen verwendet wird.

Patienten mit Aphasie vs. Sprachgesunde

Die Ergebnisse der vorliegenden Studie unterstützen die Anwendung des traditionell verwendeten Maßes (*Anzahl korrekter Nennungen*) für die Evaluation von Wortgenerierungsleistungen bei Patienten mit Aphasie und bestätigen darüber hinaus die Anwendbarkeit von *machine learning*-Techniken auf Wortgenerierungsdaten. Die Variable *Anzahl korrekter Nennungen* schien ein adäquates Maß zu sein, um eine Gruppe von Patienten mit leichter bis mittelschwerer Aphasie von Sprachgesunden zu dissoziieren; zusätzliche Variablen verringerten die Klassifikationsgenauigkeit. Darüber hinaus wurde die Anwendung der *Anzahl korrekter Nennungen* bei der Klassifikation von Patienten mit Aphasie und Sprachgesunden durch die Ergebnisse der Variablenanalyse unterstützt (vgl. Tabelle 4.13). Die Variablen $nr_resp_{L_c}$ und $nr_resp_{L_a}$ waren entsprechend ihrer χ^2 -Werte auf den ersten beiden Rängen gelistet, was zeigt, dass diese Variablen für die Klassifikation am besten geeignet waren. Bezüglich der Baseline-Strategie (BL) zeigen die Tabellen 4.3 bis 4.5, dass die Wechsel-Aufgaben die besten Ergebnisse erzielten, gefolgt von der semantischen und den beiden lexikalischen Aufgaben. Vor allem die komplexeren Wechsel-Aufgaben schienen gut für die Dissoziation zwischen Patienten mit Aphasie und Sprachgesunden geeignet zu sein.

Patienten mit KKD vs. Sprachgesunde

Interessanterweise ergaben die Ergebnisse der Klassifikation von Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden, dass Patienten mit kommunikativ-kognitiven Defiziten am besten klassifiziert werden können, wenn traditionelle Maße unberücksichtigt bleiben. Wie bereits erwähnt, ist die Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden eine Herausforderung, weil traditionelle Maße und Verfahren aufgrund der vergleichsweise geringen Beeinträchtigungen nicht immer reliabel zwischen beiden Gruppen unterscheiden können (Turkstra et al., 2005a,b). Die beschriebenen Ergebnisse bestätigen dies, da sie zeigen,

dass Patienten mit kommunikativ-kognitiven Defiziten am besten anhand anderer als der traditionell verwendeten Maße klassifiziert werden können. Darüber hinaus lieferte das Ranking aus der Variablenanalyse erste Hinweise auf die Variablen, die eventuell eine bessere Klassifikation von Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden ermöglichen können (Tabelle 4.15). Die Variablen *shortest_break_{L_c}* und *shortest_break_{L_a}* waren entsprechend ihrer χ^2 -Werte als Beste gelistet. Dies könnte darauf hinweisen, dass zeitliche Maße besonders geeignet sind, um zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden zu dissoziieren. Andere Studien, die ebenfalls zeitliche Charakteristika analysierten berichten ebenfalls von einer beeinträchtigten Leistung bei Patienten mit kommunikativ-kognitiven Defiziten (Barrow et al., 2003; Barrow, Hough, Rastatter, Walker, Holbert & Rotondo, 2006).

Hohe χ^2 -Werte wurden vor allem für die Wechsel-Aufgaben erzielt, mit höheren Werten für die semantische im Vergleich zur lexikalischen Wechsel-Aufgabe. Bei Patienten mit kommunikativ-kognitiven Defiziten sind Frontalhirn- bzw. rechtshemisphärische Läsionen mit Beeinträchtigungen der Exekutivfunktionen assoziiert (vgl. Abschnitt 2.4.2). Die alternierende Produktion von Begriffen aus zwei verschiedenen Kategorien erfordert, neben anderen kognitiv-exekutiven Funktionen, die Inhibition der einen Kategorie, um Wörter der nächsten Kategorie zu produzieren. Daher scheint es plausibel, dass diese Aufgaben sehr gut zwischen Sprachgesunden und Patienten mit kommunikativ-kognitiven Defiziten dissoziieren.

Zukünftige Forschung in diesem Bereich sollte sich auf eine noch detailliertere Analyse der verschiedenen Variablen fokussieren, eventuell mit einem besonderen Augenmerk auf die Analyse zeitlicher Charakteristika. Das Ziel sollte dabei sein Variablen zu identifizieren, die eine reliable Dissoziation zwischen Sprachgesunden und Patienten mit kommunikativ-kognitiven Defiziten erlauben. So könnten Variablen beispielsweise einen Vergleich der Anzahl der Nennungen in verschiedenen Zeitintervallen beinhalten (z.B. *Anzahl der Nennungen* in den ersten 15 Sekunden im Vergleich zu den letzten 15 Sekunden) oder Pausenzeiten vor *Cluster*-Wechseln einbeziehen (Mayr, 2002; Troyer & Moscovitch, 2006). Insbesondere die Analyse von auf *Cluster* und *Switches* bezogene zeitliche Charakteristika scheint besonders vielversprechend. Die qualitative und quantitative Analyse von *Cluster* und *Switches* (ohne Berücksichtigung zeitlicher Aspekte) ist ein häufig beschriebenes klinisches Maß, die Auswertung der *Anzahl korrekter Nennungen* ergänzend. Es ist darüber hinaus bekannt, dass im Rahmen von Wortgenerierungsaufgaben Nennungen in sogenannten "spurts" (d.h. stoßweise) produziert werden. Wortproduktionen innerhalb einer Subkategorie folgen relativ rasch aufeinander; wird zu einer neuen Subkategorie gewechselt (*Switching*) entsteht üblicherweise eine im Vergleich längere Pause (Troyer & Moscovitch, 2006; Unsworth et al., 2011). Durch die Analyse von Pausenzeiten zwischen Sub-

kategorien könnten sich weitere Hinweise zur Relevanz zeitlicher Charakteristika bei der Evaluation von Wortgenerierungsleistungen ergeben. Auch insgesamt kann die Betrachtung zusätzlicher Variablen weitere Erkenntnisse für eine sichere Identifizierung und reliable Klassifikation von Patienten mit kommunikativ-kognitiven Defiziten liefern.

Wenn erst einmal hinreichend automatisiert, erfordern die Anwendung von *machine learning*-Techniken oder die Variablenanalyse nur geringfügigen zusätzlichen zeitlichen und finanziellen Aufwand im Vergleich zur klinischen Auswertung. Ihre Aussagekraft und Klassifikationsgenauigkeit übertreffen die klinische Evaluation jedoch. Die beschriebenen Analysen könnten zukünftig ein sinnvolles Werkzeug sein, um Sprach- und/oder Kommunikationsstörungen (auch subtile Beeinträchtigungen) nach Hirnschädigung zu untersuchen. Dies insbesondere bei solchen Störungsbildern, die mit Standardprozeduren und traditionellen klinischen Analysen nicht reliabel voneinander oder von Sprachgesunden abgegrenzt werden können.

Patienten mit Aphasie vs. Patienten mit KKD

Für die Dissoziation zwischen Patienten mit Aphasie und Patienten mit kommunikativ-kognitiven Defiziten wurden die besten Ergebnisse mit der Baseline-Strategie *Anzahl korrekte Nennungen* erzielt. Die beiden Variablen $nr_resp_{L_c}$ und $nr_resp_{L_a}$ waren auch bei der Variablenanalyse als beste gelistet (vgl. Tabelle 4.14). Dies liefert Hinweise darauf, dass ein Maß wie die *Anzahl korrekter Nennungen* reliabel zu sein scheint, um Patienten mit Aphasie von anderen Patientengruppen oder Sprachgesunden zu unterscheiden. Die Versuchsteilnehmer in der Gruppe der Aphasie-Patienten waren im Hinblick auf den Schweregrad der Aphasie (leicht bis mittelschwer) relativ heterogen. Weitere Untersuchungen, die zwischen Schweregraden und Aphasiesyndromen unterscheiden und insbesondere Untersuchungen von Patienten mit Restaphasie sind erforderlich, um weitergehende Aussagen machen zu können.

Basierend auf einem allgemeinen Konsens (Huber et al., 1983), werden Aphasien gewöhnlich entsprechend der Flüssigkeit der Sprachproduktion (flüssig vs. nicht-flüssig), der Aphasiesyndrome (z.B. Amnestische Aphasie, Broca Aphasie, Wernicke Aphasie, Globale Aphasie) oder des Schweregrades (leicht, mittelschwer, schwer) in verschiedene Unterformen eingeteilt. Eine Evaluation charakteristischer Merkmale für die Dissoziation verschiedener Unterformen durch Leistungen bei der Wortgenerierung ist für die zukünftige Forschung von Relevanz. Demgegenüber ist bei Patienten mit kommunikativ-kognitiven Defiziten die Existenz von Unterformen noch Gegenstand der Diskussion (Crowe, 1992; Heidler, 2009; Martin & McDonald,

2003). Zukünftige Studien könnten sich daher auf mögliche Unterformen kommunikativ-kognitiver Defizite fokussieren, insbesondere zeitliche Charakteristika berücksichtigend, indem sie den relativen Einfluss bestimmter Merkmale auf die Klassifikation von kommunikativ-kognitiven Defiziten und möglichen Unterformen untersuchen. Hier könnten dann auch andere Algorithmen des *machine learning*, wie zum Beispiel das *data mining* Anwendung finden.

4.4.1 Fazit

In dieser Studie wurden eine Reihe zusätzlicher Variablen für die Evaluation von Wortgenerierungsaufgaben vorgestellt. Unter der Verwendung von *machine learning*-Techniken wurden diese Variablen hinsichtlich ihres Einflusses auf die Klassifikationsgenauigkeit bei Sprachgesunden und zwei Patientengruppen, Patienten mit Aphasie und Patienten mit kommunikativ-kognitiven Defiziten, untersucht. Bei der Klassifikation von Patienten mit Aphasie und Sprachgesunden unterstützen die beschriebenen Ergebnisse den Vorteil des traditionell angewendeten Maßes *Anzahl korrekter Nennungen* für die Evaluation von Wortgenerierungsaufgaben. Im Gegensatz dazu sollten bei der Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden andere Maße als die *Anzahl korrekter Nennungen* angewendet werden, um Wortgenerierungsleistungen zu evaluieren. Eine besonders interessante Variable scheint dabei die Länge der kürzesten Zeitspanne zwischen den Startzeitpunkten zweier konsekutiver Nennungen zu sein.

Die beschriebenen Resultate haben das Potenzial, langfristig zu einer verbesserten klinischen Datenanalyse zu führen. Die Identifizierung geeigneter Wortgenerierungsaufgaben und spezifischer Auswertungsparameter kann dabei helfen, die Dissoziation zwischen den beschriebenen Patientengruppen und Sprachgesunden zu faszilitieren. Die Ergebnisse können damit einen Einfluss auf bestehende klinische Herangehensweisen haben und dazu beitragen die Herausforderung zu meistern, reliabel zwischen Sprachgesunden und Patienten mit Sprach- und/oder Kommunikationsstörungen dissoziieren zu können.

Diese Studie hat gezeigt, dass *machine learning*-Techniken eine geeignete Methode zur Identifizierung spezifischer Störungsbilder sein können. Im nächsten Kapitel folgt nun eine weitere Studie, die sich mit der Anwendung von *machine learning*-Techniken auf Wortgenerierungsdaten befasst. Auf Basis eines explorativen Ansatzes wurden in Kooperation mit Kollegen aus der Informatik verschiedene Möglichkeiten der Evaluation von Wortgenerierungsleistungen mit *machine learning*-Techniken zur Klassifikation von Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden miteinander verglichen.

Kapitel 5

Exploration einer Methodologie zur Evaluation von Wortgenerierungsleistungen mit *machine* *learning*-Techniken

5.1 Einleitung & Fragestellung

Kommunikativ-kognitive Beeinträchtigungen werden als Folge kognitiver Defizite beschrieben, welche sich auf höhere sprachliche Funktionen und kommunikative Leistungen auswirken und häufig nach rechtshemisphärischen und Frontalhirnläsionen verschiedenster Ätiologien auftreten (vgl. hierzu auch Abschnitt 2.4.2 oder siehe Coelho & DeRuyter, 1996; Cummings, 2007; Lê et al. 2011). Während sprachstrukturelle Leistungen bei Patienten mit kommunikativ-kognitiven Defiziten als erhalten gelten, wird die Anwendung von Sprache in der sozialen Interaktion als auffällig beschrieben (Martin & McDonald, 2003). Insbesondere Beeinträchtigungen in den Exekutivfunktionen werden als Ursache kommunikativ-kognitiver Defizite gesehen (Martin & McDonald, 2003). Eine Möglichkeit der Evaluation von Leistungen in den Exekutivfunktionen besteht in der Überprüfung der Wortgenerierungsfähigkeit. Diese wird bei Patienten mit kommunikativ-kognitiven Defiziten als beeinträchtigt beschrieben (Bittner & Crowe, 2006; Kavé et al., 2011; Martin & McDonald, 2003). Die nun folgende Studie untersucht die Wortgenerierungsleistung von Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden unter Berücksichtigung verschiedener Methodologien.

Wie in Kapitel 2 (vgl. Abschnitt 2.6.1) beschrieben und auch durch die vorausgehende Studie in Kapitel 4 ersichtlich, stellt der Einsatz von *machine learning*-Techniken einen Erfolg versprechenden Ansatz dar, um die Diagnosestellung bei verschiedenen klinischen Populationen langfristig zu vereinfachen. So berichten verschiedene Autoren die Anwendbarkeit von *machine learning*-Techniken bei der Identifizierung sowie bei der Differentialdiagnose aphasischer Sprachstörungen, psychiatrischer Erkrankungen oder verschiedener demenzieller Erkrankungen etc. auf Basis linguistischer oder kognitiver Maße sowie anhand von Bildgebungsbefunden (Costafreda et al., 2011; Fraser et al., 2012; Järvelin, 2008; Järvelin & Juhola, 2011; Maier et al., 2011; Quintana et al., 2012; Wilson et al., 2009). Zu erwähnen sind hier auch die Arbeiten der Gruppe um Brian Roark (Lehr, Prud'hommeaux, Shafran, & Roark, 2012; Prud'hommeaux & Roark, 2011; Prud'hommeaux, Mitchell, & Roark, 2011; Roark, Mitchell, & Hollingshead, 2007; Roark, Mitchell, Hosom, Hollingshead, & Kaye, 2011), dabei insbesondere die Befunde von Lehr und Kollegen (2012). Die Autoren konnten zeigen, dass mit der Anwendung von *machine learning*-Techniken auf neuropsychologischen Testdaten ein reliables, automatisiertes Screening zur Identifizierung leichter kognitiver Beeinträchtigungen in großen Kohorten möglich ist.

Die folgende Studie stellt eine Erweiterung der in Kapitel 4 beschriebenen Analysen dar. Im vorausgehenden Kapitel wurde eine Studie vorgestellt bei der *machine learning*-Techniken auf Wortgenerierungsdaten angewendet wurden, um verschiedene zusätzliche Variablen zu evaluieren, mit dem Ziel Patienten mit Aphasie, Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunde zu klassifizieren. Während die Analysen die Evaluation der *Anzahl korrekter Nennungen* als geeignetes Maß für die Dissoziation zwischen Patienten mit Aphasie und Sprachgesunden bestätigten, erzielten die zusätzlich untersuchten Variablen bessere Ergebnisse bei der Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden im Vergleich zu traditionell verwendeten klinischen Maßen. Demnach scheint bei der Evaluation von Wortgenerierungsleistungen bei Patienten mit kommunikativ-kognitiven Defiziten die Anwendung anderer Variablen als die *Anzahl korrekter Nennungen* besser geeignet zu sein.

Mit der nun folgenden Studie wird die Möglichkeit der Entwicklung einer Evaluationsmethode zur Diagnose kommunikativ-kognitiver Defizite auf Basis von Wortgenerierungsdaten exploriert. Die Entwicklung eines Verfahrens zur Diagnose kommunikativ-kognitiver Defizite ist insbesondere deshalb relevant, da bislang kein allgemein akzeptiertes und reliables Testverfahren existiert (Turkstra et al., 2005a,b).

Als Basis für die Exploration der Entwicklung eines solchen Verfahrens werden geeignete Variablen aus der vorausgehenden Studie (vgl. Kapitel 4) eingehender, sowie einige neue Variablen evaluiert. Mit diesen Variablen und

einem Referenzdatenset, welches aus Wortgenerierungsdaten aus vier spezifischen Aufgaben besteht, werden drei verschiedene Experimente durchgeführt. Im ersten Experiment werden individuelle Variablen sowohl mit *machine learning*-Techniken, das heißt, Training und Evaluation mit Klassifikatoren, als auch mit klinisch-statistischen Analysen evaluiert. Die Effizienz individueller Variablen wird gemittelt über die vier Wortgenerierungsaufgaben untersucht und dem traditionell bei der klinischen Auswertung verwendeten Leistungsmaß *Anzahl korrekter Nennungen* gegenübergestellt. Da die verschiedenen Wortgenerierungsaufgaben zum Teil unterschiedliche kognitive Funktionen abtesten (Basso et al., 1997; Nutter-Upham et al., 2008; Troyer et al., 1997), wird darüber hinaus die Effizienz individueller Variablen in den einzelnen Wortgenerierungsaufgaben separat überprüft. Im zweiten Experiment wird der Frage nachgegangen, ob Klassifikatoren, die auf verschiedenen Variablenkombinationen trainiert werden, besser für die Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden geeignet sind als Klassifikatoren, die auf einzelnen Variablen trainiert werden. Neben der Klassifikationsgenauigkeit (*accuracy*), welche in der vorausgehenden Studie (vgl. Kapitel 4) als Evaluationsmaß diente, werden hier zwei weitere Maße herangezogen: der positive Vorhersagewert (*Precision*) und die Sensitivität (*Recall*). Diese beiden Maße erlauben zusätzlich gruppenspezifische Aussagen über die Effizienz der ermittelten Variablenkombinationen im Hinblick auf die Anwendbarkeit bei der Diagnosestellung. Sie stellen qualitative Kennwerte diagnostischer Verfahren dar und werden dementsprechend als wichtiges Maß der Güte eines Tests gesehen (vgl. Bortz & Lienert, 2008; Keil & Kaszniak, 2002). Im dritten Experiment wird exploriert, ob das Trainieren von Klassifikatoren auf einzelnen Wortgenerierungsaufgaben mit aufgabenspezifischen Variablenkombinationen und ein anschließendes Zusammenführen der Ergebnisse zu einer Gesamtentscheidung, das heißt die Anwendung eines Meta-Klassifikators, die Klassifikationsgenauigkeit weiter verbessern kann.

Im Folgenden wird zunächst die verwendete Methode, einschließlich einer Beschreibung des verwendeten Datensatzes, der berücksichtigten Variablen sowie der Evaluation mit *machine learning*-Techniken, dargestellt. Anschließend werden die drei auf dieser Basis durchgeführten Experimente präsentiert. Einem abschließenden Fazit geht eine allgemeine Diskussion der Ergebnisse sowie limitierender Aspekte und möglicher Erweiterungen voraus. Auszüge der folgenden Studie sind beschrieben in Thiele, Gaspers, Cimiano, Foltz, & Stenneken (under revision).

5.2 Methode (allgemein)

Einen Überblick über den Ablauf, beginnend mit der Datenerhebung bis hin zur Durchführung der einzelnen Experimente, gibt Abbildung 5.1. Zunächst wurden mit verschiedenen Wortgenerierungsaufgaben Daten von kommunikativ-kognitiv beeinträchtigten Patienten und Sprachgesunden erhoben. Für jede Aufgabe ergab sich eine Liste mit Nennungen der Versuchsteilnehmer. Die Nennungen wurden transkribiert und entlang einer Zeitachse mittels ELAN (Max Planck Institute for Psycholinguistics, 2011; Sloetjes & Wittenburg, 2008) annotiert. Die resultierende Liste der Nennungen wurde verwendet um Variablenvektoren zu extrahieren. Mit diesen Vektoren wurden anschließend verschiedene Experimente durchgeführt, um zu explorieren wie eine Evaluationsmethode für die Diagnose kommunikativ-kognitiver Defizite auf Basis von Wortgenerierungsaufgaben konstruiert sein könnte. Es folgt an dieser Stelle eine detaillierte Beschreibung der einzelnen Schritte.

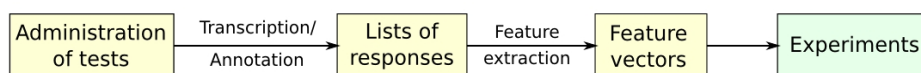


Abbildung 5.1: Workflow: Datenerhebung (*Administration of tests*), Transkription/Annotation, Liste der Nennungen (*List of responses*), Extraktion der Variablen (*Feature extraction*), Variablenvektoren (*Feature vectors*), Experimente; aus Thiele et al. (under revision, S. 5)

5.2.1 Versuchsteilnehmer

Insgesamt wurden Wortgenerierungsdaten von 28 Versuchsteilnehmern erhoben. Diese gliederten sich in eine Gruppe von Patienten mit kommunikativ-kognitiven Defiziten und eine Kontrollgruppe. Ein Teil der Daten (Daten von 13 Patienten mit kommunikativ-kognitiven Defiziten und einem Teilnehmer der Kontrollgruppe) wurde im Rahmen einer größeren Studie erhoben und bereits teilweise für Analysen verwendet, die in Gaspers et al. (2012) beschrieben wurden. Alle Teilnehmer waren monolinguale, deutsche Muttersprachler mit normalem oder korrigiertem Seh- und Hörvermögen. Die Datenerhebung erfolgte in Übereinstimmung mit den 1964 in der Helsinki Deklaration festgelegten ethischen Standards. Alle Teilnehmer gaben nach erfolgter Aufklärung ihre Einwilligung zur Teilnahme an der Studie.

Patienten mit kommunikativ-kognitiven Defiziten (KKD). Die erste Gruppe bestand aus 14 Patienten mit kommunikativ-kognitiven Defiziten (alle männlich) im Alter von 18 bis 58 Jahren ($M = 40.57$, $SD = 12.76$). Neurologische Befunde der Ätiologie umfassten Schädelhirntraumata ($n = 11$), zerebralen Tumor ($n = 1$) sowie ischämischen ($n = 1$) und hämorrhagischen ($n = 1$) Schlaganfällen.

gischen ($n = 1$) Insult mit einer Zeit post onset von 1 bis 173 Monate ($M = 34.07$, $SD = 55.25$). Die Diagnose einer kommunikativ-kognitiven Störung erfolgte auf Basis manifester klinischer Symptome, wie z.B. befremdliche Ausdrucksweise, semantische Vagheit und/oder umständliche Formulierungen, stark ausschweifende und/oder sozial nicht akzeptable Äußerungen etc. (vgl. Lê et al., 2011; Martin & McDonald, 2003; Prigatano, 1986)³¹. Die Diagnose eines KKD wurde gestellt, wenn klinisch Symptome einer kommunikativ-kognitiven Beeinträchtigung von mindestens zwei erfahrenen Beurteilern (z.B. Sprachtherapeut, Arzt) unabhängig voneinander beschrieben wurden. Um auszuschließen, dass sprachsystematische Störungen Ursache der beobachteten Symptome sind, wurden alle Versuchsteilnehmer dieser Gruppe zusätzlich mit dem Aachener Aphasia Screening (*Token Test* und Untertest *Schriftsprache* aus dem AAT, Huber et al., 1983) vorab als nicht-aphasisch eingestuft. Der Bildungsgrad betrug im Mittel 11.07 Schuljahre ($SD = 1.69$, $Min = 9$, $Max = 14$).

Sprachgesunde. Die zweite Gruppe bestand aus 14 neurologisch gesunden Kontrollprobanden (alle männlich) im Alter von 22 bis 61 Jahren ($M = 41.29$, $SD = 12.41$) und mit einem mittleren Bildungsgrad von 11.21 Schuljahren ($SD = 1.72$, $Min = 8$, $Max = 14$). Anamnestisch bestand bei keinem Versuchsteilnehmer aus dieser Gruppe ein Hinweis auf das Vorliegen einer neurologischen oder psychiatrischen Erkrankung.

Beide Gruppen unterschieden sich nicht signifikant in Bezug auf Alter und Bildungsgrad (t -Test für unabhängige Stichproben: Alter: $t(26) = -0.150$, $p = .882$; Bildungsgrad: $t(26) = -0.222$, $p = .826$). Die Analysen wurden mit der Statistiksoftware PASW Statistics 18 (2009) gerechnet.

5.2.2 Material, Durchführung & Auswertung

Im Folgenden werden sowohl Materialien und Durchführung der verwendeten Wortgenerierungsaufgaben, als auch die bei der Evaluation berücksichtigten Variablen beschrieben.

Wortgenerierungsaufgaben

Um mit der Aufgabe vertraut zu werden, führten alle Versuchsteilnehmer zunächst eine lexikalische Aufgabe (LEX_A) als Übungsaufgabe durch. Vier weitere Wortgenerierungsaufgaben wurden für die anschließende Evaluation verwendet. Die Versuchsteilnehmer führten eine simple lexikalische Aufgabe, eine simple semantische Aufgabe, eine lexikalische Wechsel-Aufgabe so-

³¹Diese Herangehensweise wurde gewählt, da bislang im deutschsprachigen Raum kein allgemein akzeptiertes, standardisiertes und normiertes Verfahren zur Diagnose kommunikativ-kognitiver Defizite existiert.

wie eine semantische Wechsel-Aufgabe durch. Alle Aufgaben (bis auf die Übungsaufgabe) sind Teil des Regensburger Wortflüssigkeitstest (Aschenbrenner et al., 2000). Die Aufgabe der Versuchsteilnehmer bestand darin, unter Beachtung der jeweils vorgegebenen Regeln, innerhalb einer Zeit von 60 Sekunden so viele Wörter wie möglich zu produzieren. Alle Aufgaben wurden stets in derselben Reihenfolge präsentiert. Die Anweisung erfolgte entsprechend der Vorgaben für lexikalische, semantische und Wechsel-Aufgaben aus dem Regensburger Wortflüssigkeitstest (Aschenbrenner et al., 2000).

Bei der lexikalischen Wortgenerierungsaufgabe sollten die Versuchsteilnehmer Wörter mit dem Anfangsbuchstaben *M* produzieren (z.B. *Milch, mager, messen, Miete etc.* – LEX_M). Die semantische Aufgabe verlangte das Generieren von Wörtern aus der Kategorie *Tiere* (z.B. *Hund, Katze, Pferd, Tiger etc.* – SEM_Tiere). Bei der lexikalischen Wechsel-Aufgabe sollten alternierend Wörter mit den Anfangsbuchstaben *H* und *T* (z.B. *Hund, Tomate, hässlich, Tante etc.* – LEX_H/T) produziert werden, während die semantische Wechsel-Aufgabe die alternierende Produktion von Wörtern aus den Kategorien *Kleidung* und *Blumen* (z.B. *Hose, Rose, Kleid, Nelke etc.* – SEM_Kleidung/Blumen) erforderte.

Nach erfolgter Datenerhebung wurden die Nennungen der Versuchsteilnehmer in den einzelnen Aufgaben manuell in ELAN (Max Planck Institute for Psycholinguistics, 2011; Sloetjes & Wittenburg, 2008) annotiert. Für jeden Versuchsteilnehmer und jede Wortgenerierungsaufgabe ergab sich daraus eine spezifische Wortliste.

Variablen

Es wurden Variablen ausgewählt, die für eine Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden Erfolg versprechend waren. Dazu zählten sowohl als vielversprechend identifizierte Variablen der vorausgehenden Studie (vgl. Kapitel 4) als auch einige neue Variablen. Diese wurden aus den gewonnenen Wortgenerierungsdaten extrahiert, das heißt aus der Liste der Nennungen der Versuchsteilnehmer. Die Nennungen der Versuchsteilnehmer wurden entsprechend der im Regensburger Wortflüssigkeitstest (Aschenbrenner et al., 2000) formulierten Regeln hinsichtlich ihrer Korrektheit bewertet; alle inkorrekten Nennungen wurden von der weiteren Analyse ausgeschlossen.

Im Folgenden wird ein Überblick über die untersuchten Variablen gegeben. Jede Variable wurde aus der Liste der korrekten Nennungen einer Wortgenerierungsaufgabe extrahiert.

- **Baseline** (*Anzahl korrekter Nennungen*): Als Baseline wurde eine Variable verwendet, die traditionell in der klinischen Praxis bei der Evaluation von Wortgenerierungsleistungen verwendet wird. Entsprechend der Regeln im Regensburger Wortflüssigkeitstest (vgl. Aschenbrenner et al., 2000) wurde die Anzahl der korrekten Nennungen berechnet. Diese ergab die Baseline-Variable *nr_responses*.
- **Zeitintervalle** (*Anzahl korrekter Nennungen in unterschiedlichen Zeitintervallen*): Die Anzahl korrekter Nennungen kann sich über die Zeit verändern (Adams et al., 1989; Bittner & Crowe, 2006; Hall et al., 2011; Raboutet et al., 2010), auch in unterschiedlicher Art und Weise für Patienten mit Sprach-/Kommunikationsstörungen und Sprachgesunde. So könnte es sein, dass sich auch für Patienten mit KKD zeigt, dass sie zu Beginn noch viele, jedoch zum Ende der Aufgabe nur noch wenige Nennungen produzieren (vgl. Hall und Kollegen, 2011, für Wortgenerierungsleistungen von kognitiv-beinträchtigten Älteren). Daher wurde die Anzahl korrekter Nennungen in zwei Intervallen berechnet: i) in den ersten 30 Sekunden (*first_interval*) der Aufgabe und ii) zwischen Sekunde 31 und 60 (*second_interval*).
- **Cluster und Switches** (*Clustergröße und Anzahl von Switches*): Nach Troyer und Kollegen (1997) sind *Cluster* (Anzahl von Nennungen in Subkategorien) und *Switches* (Anzahl der Wechsel zwischen Subkategorien) wichtige Komponenten der Wortgenerierungsfähigkeit, die für eine optimale Leistung erforderlich sind. Für die simple lexikalische (LEX_M) und die simple semantische (SEM_Tiere) Aufgabe wurden daher die Anzahl der *Switches* sowie die durchschnittliche *Clustergröße* ermittelt. *Clustergröße* und *Switches* wurden per Hand annotiert, entsprechend der Regeln in Troyer et al. (1997). Daraus ergaben sich die Merkmale *nr_switches* und *avg_clusters*.
- **Zeitliche Aspekte** (*Pausendauer*): Veränderungen zeitlicher Charakteristika sind für verschiedene klinische Populationen beschrieben worden (vgl. Gaspers et al., 2012; Koren et al., 2005; Rosen et al., 2005). Zur Analyse der zeitlichen Variation bei der Produktion von Nennungen wurde die Dauer aller Pausen berechnet, wobei sich die Pausendauer als Zeit zwischen dem Ende einer Nennung und dem Beginn der folgenden Nennung berechnete. Beginn und Ende der Nennungen waren in ELAN (Max Planck Institute for Psycholinguistics, 2011; Sloetjes & Wittenburg, 2008) annotiert und konnten entsprechend ausgelesen werden. Als zeitliche Variablen wurden berechnet: die durchschnittliche Pausendauer (*avg_break*), die längste Pause (*longest_break*) und die kürzeste Pause (*shortest_break*) sowie auch die Pausenvarianz (*var_break*).

- **Lexikalische Aspekte** (*Wortfrequenz (gesprochen) und Anzahl von Google Hits*): Einige Studien haben den Einfluss der Wortfrequenz auf die Wortproduktion bei Patienten mit neurogenen Störungen beschrieben (vgl. Ellis et al., 1983; Nickels & Howard, 1995). Darüber hinaus berichtete Crowe (1998) bei Sprachgesunden, dass sich Wortproduktionsraten und Wortfrequenz verringern, wenn die für die Durchführung einer Wortgenerierungsaufgabe zur Verfügung stehende Zeit voranschreitet. Aus diesem Grund wurden auch Wortfrequenz-Variablen berücksichtigt. Wortfrequenzen gesprochener Sprache wurden mit dem von Aichert, Marquardt und Ziegler (unveröff.) beschriebenen Computerprogramm für jede korrekte Nennung ermittelt. Wörter, die nicht in der Datenbasis des Programms enthalten waren, wurden als fehlende Werte behandelt. Berechnet wurden die durchschnittliche Wortfrequenz (*avg_frequency*), der höchste (*highest_frequency*) und der niedrigste Wortfrequenzwert (*lowest_frequency*) sowie die Varianz der Wortfrequenzwerte (*var_frequency*). Da die Versuchsteilnehmer viele niederfrequente Wörter produzierten, die nicht in der Datenbasis des Computerprogramms enthalten waren, wurde zusätzlich das Internet zur Wortfrequenzbestimmung genutzt. Für jede korrekte Nennung wurde die Anzahl von *Google Hits*, das heißt die Anzahl der Treffer, die Google bei der Suche nach einem Wort erzielt, abgerufen. Die *Google Hits* wurden mit der deutschen Version von Google (www.google.de) am 14. November 2012 ermittelt. Anschließend wurden die durchschnittliche Anzahl von *Google Hits* (*avg_google*), der höchste Wert (*highest_google*), der niedrigste Wert (*lowest_google*) und die Varianz der Werte (*var_google*) berechnet.
- **Spezifität/semantischer Gehalt** (*Anzahl übergeordneter Kategorien in Wikipedia*): Auch die Spezifität der Nennungen der Versuchsteilnehmer wurde berücksichtigt, da Unterschiede bezüglich des semantischen Gehalts von Nennungen für Patienten mit Aphasie sowie Patienten mit nicht-aphasischen Beeinträchtigungen nach rechts-hemisphärischer Hirnschädigung beschrieben wurden (Adams et al., 1989; Grossman, 1981). Die Spezifität wurde mit Hilfe von Wikipedia für alle Nomen ermittelt.³² Zunächst wurde automatisiert die Anzahl übergeordneter Kategorien in Wikipedia für jedes genannte Nomen gezählt, das heißt die Anzahl von Klicks auf den Kategorie-Link, der von der Seite der jeweiligen Nennung bis zur Seite der "Hauptkategorie" führt. Anschließend wurden alle Daten manuell kontrolliert, um zu gewährleisten, dass für alle Nennungen die semantisch am besten passenden Verknüpfungen ermittelt wurden. Ambige Nennungen (das heißt wenn Wikipedia eine Disambiguierung einforderte, wie z.B. bei den

³²Verben und Adjektive wurden ausgeschlossen, da Wikipedia keine verwertbaren Einträge für Wörter dieser Wortklassen enthält.

verschiedenen Bedeutungen von "m/Morgen") wurden aus der Analyse der Spezifität ausgeschlossen. Berechnet wurden die durchschnittliche (*avg_specificity*), höchste (*highest_specificity*) und niedrigste (*lowest_specificity*) Spezifität, sowie auch die Varianz der Spezifität (*var_specificity*).

- **Sublexikalische Aspekte** (*Silben- und Phonemanzahl pro Nennung*): Es wurde vermutet, dass von Patienten mit Sprach- und/oder Kommunikationsbeeinträchtigung generell weniger komplexe Wörter verwendet würden, im Vergleich zu Sprachgesunden (Nickels & Howard, 2004; Pekkala, 2012). Deshalb wurden die Silben- und Phonemanzahl pro Nennung als Variablen in die Analyse einbezogen. Zur Ermittlung der Silben- und Phonemanzahl für jede korrekte Nennung wurde erneut das Computerprogramm, beschrieben in Aichert et al. (unveröff.), verwendet. Da die Versuchsteilnehmer eine Reihe von Pseudowörtern und niederfrequenten Wörtern produziert hatten, waren nicht alle Wörter in der Datenbasis des Programms enthalten. Bei allen Wörtern, die nicht in der Datenbasis enthalten waren, wurden Silben- und Phonemanzahl manuell annotiert. Anschließend wurden die durchschnittliche Silben- (*avg_syll*) und Phonemanzahl (*avg_phon*), die kleinste Anzahl von Silben (*lowest_syll*) und Phonemen (*lowest_phon*), die größte Anzahl von Silben (*highest_syll*) und Phonemen (*highest_phon*) sowie die Varianz der Anzahl von Silben (*var_syll*) und Phonemen (*var_phon*) berechnet.

5.2.3 Evaluation mit *machine learning*-Techniken

Neuere Arbeiten, die die Anwendbarkeit von Klassifikatoren bei der Dissoziation verschiedener Patientengruppen untersucht haben, weisen darauf hin, dass, abhängig von den Datensätzen und der Klassifikationsaufgabe, verschiedene Klassifikatoren die besten Ergebnisse erzielen können (Järvelin & Juhola, 2011). Das heißt, es gibt keinen universellen Klassifikator, der für alle Klassifikationsaufgaben und Datensätze gleich gut geeignet ist. Daher scheint es angeraten, den Klassifikator unter Berücksichtigung der jeweiligen Datensätze und Experimente auszuwählen.

Primäres Ziel dieser Studie war, unterschiedliche Evaluationsmethoden zu explorieren und zu vergleichen und nicht die Identifizierung des leistungsstärksten Klassifikators für diese Evaluation. Die Wahl des Klassifikators erfolgte daher ausschließlich auf Basis von Ergebnissen einer früheren Studie. In Gaspers et al. (2012) dissoziierte der 1NN Klassifikator am besten zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden und wurde daher auch für die Experimente in dieser Studie ausgewählt. Obwohl in der früheren Studie ein etwas anderes Datenset verwendet wurde,

sind die beiden Studien insofern vergleichbar, als in beiden nur eine relativ geringe Menge an Wortgenerierungsdaten von Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden verwendet wurde bzw. wird. Der 1NN Klassifikator war daher sehr wahrscheinlich auch für die folgenden Experimente dieser Studie eine gute Wahl.

Zusammen mit dem 1NN Klassifikator und den Variablenvektoren aus den verschiedenen Wortgenerierungsaufgaben, wurden verschiedene Experimente durchgeführt mit dem Ziel zu untersuchen, wie eine Evaluationsmethode für die Diagnose kommunikativ-kognitiver Defizite auf Basis von Wortgenerierungsdaten konstruiert sein könnte. Während dieses Prozesses wurden verschiedene 1NN Klassifikatoren (z.B. bei der Verwendung verschiedener Variablen und Kombinationen von Wortgenerierungsaufgaben) für die Klassifikation von 14 Patienten mit kommunikativ-kognitiven Defiziten und 14 sprachgesunden Kontrollprobanden gebildet und verglichen. Die Performanz des Klassifikators wurde in Form der *accuracy* gemessen. Die einzelnen Experimente werden in den folgenden Abschnitten vorgestellt. Sie wurden mit der *Weka Data Mining Software* (Version 3.6.2.; Hall et al., 2009) durchgeführt. An dieser Stelle könnte man den Einsatz verschiedener Kreuzvalidierungsmethoden, wie etwa der *leave-one-out*-Kreuzvalidierung (vgl. Maier, Höning, Steidl, Nöth, Horndasch, Sauerhöfer, Kratz, & Moll, 2011) oder einer *5x2-fach*-Kreuzvalidierung (vgl. Dietterich, 1998), diskutieren. Die im Folgenden beschriebenen Experimente wurden auf Basis einer 5-fach-Kreuzvalidierung durchgeführt.

5.3 Experiment 1

Mit dem ersten Experiment wurde untersucht, wie effizient die einzelnen Variablen, die im vorausgehenden Abschnitt eingeführt wurden, für die Evaluation von Wortgenerierungsleistungen sind. Es sollte ermittelt werden, welche Variablen am besten zur Analyse von Wortgenerierungsdaten geeignet sind, wenn das Ziel darin besteht, Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunde mittels *machine learning*-Techniken voneinander zu dissoziieren. Darüber hinaus war von Interesse, die Leistung der Variablen mit dem traditionell in der klinischen Praxis verwendeten Maß zu vergleichen.

5.3.1 Experiment 1a

Die verschiedenen Wortgenerierungsaufgaben (d.h. lexikalische, semantische und Wechsel-Aufgaben) werden traditionell unter Verwendung desselben Maßes, das heißt anhand der *Anzahl korrekter Nennungen*, ausgewertet. Mit diesem ersten Experiment sollte ermittelt werden, welche Variablen am besten geeignet waren, um alle Wortgenerierungsaufgaben zu evaluieren, das

heißt wenn dieselbe Variable auf alle Wortgenerierungsaufgaben angewendet und über alle Aufgaben gemittelt wurde. Abschließend wurde die Klassifikationsgenauigkeit³³ der Variablen verglichen.

Methoden

Auf den Daten jeder Wortgenerierungsaufgabe wurde für jede einzelne Variable ein 1NN Klassifikator trainiert. Es gab demnach vier Klassifikationsaufgaben, eine für jede Wortgenerierungsaufgabe. Die für jede Variable³⁴ ermittelte *accuracy* wurde dann über alle vier Wortgenerierungsaufgaben gemittelt.

Ergebnisse & Diskussion

Tabelle 5.1 zeigt eine Liste der zehn Variablen, die die beste *accuracy* erzielten und gibt darüber hinaus die Klassifikationsgenauigkeit der Baseline-Variable an. Gemittelt über alle Wortgenerierungsaufgaben erzielte die Variable *lowest_phon* die besten Ergebnisse mit einer *accuracy* in Höhe von 63.39%. Dies stellte eine Verbesserung im Vergleich zur Baseline von ungefähr 24% dar. Die Variable *shortest_break* erzielte ähnlich gute Ergebnisse mit einer *accuracy* in Höhe von 62.5% und einer Verbesserung im Vergleich zur Baseline von etwa 23%.

Tabelle 5.1: *Liste der besten zehn Variablen (gemittelt über alle Aufgaben) - Experiment 1a*

Variablen	Accuracy
<i>lowest_phon</i>	63.39
<i>shortest_break</i>	62.5
<i>highest_phon</i>	58.04
<i>avg_phon</i>	56.25
<i>first_interval, avg_break</i>	53.57
<i>avg_google, var_syll</i>	52.68
<i>var_specifcity, highest_syll</i>	51.79
<i>nr_responses</i>	39.29

Jede Variable auf der Liste der besten zehn erzielte eine *accuracy* oberhalb der Zufallswahrscheinlichkeit, welche 50% beträgt, da mit derselben Anzahl von Patienten mit KKD und Sprachgesunden klassifiziert wurde. Einige Variablen erzielten jedoch nur geringfügig bessere Ergebnisse im Vergleich zur Zufallswahrscheinlichkeit. Zu beachten ist jedoch, dass jede der evaluierten Variablen (nicht nur die besten zehn) eine bessere *accuracy* im

³³Zur Bezeichnung der Güte des Klassifikationsergebnisses werden in dieser Studie sowohl die englische Bezeichnung *accuracy* als auch die entsprechende deutsche Übersetzung *Klassifikationsgenauigkeit* verwendet.

³⁴Außer für die Variablen *nr_switches* und *avg_clusters*, welche nur bei den Aufgaben LEX_M und SEM_Tiere vorliegen.

Vergleich zur Baseline-Variable *nr_responses* erzielte. Mit einer *accuracy* in Höhe von von 39.29% war die Klassifikationsgenauigkeit der Baseline-Variable schlechter als bei einer zufälligen Klassifikation.

Die Ergebnisse weisen darauf hin, dass andere Variablen als die Baseline-Variable *nr_responses* in Betracht gezogen werden sollten, um Patienten mit kommunikativ-kognitiven Defiziten von Sprachgesunden zu dissoziieren. Dies vor allem, wenn die Abgrenzung auf Basis von Wortgenerierungsdaten und mit *machine learning*-Techniken vorgenommen wird. Es scheint, dass viele der anderen, zusätzlich evaluierten Variablen besser geeignet sein könnten, um zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden zu dissoziieren. Die Baseline-Variable *nr_responses* schien die Klassifikation eher zu erschweren als zu fazitätieren, was sich dadurch zeigte, dass sie nicht nur die geringste *accuracy* erzielte, sondern dabei auch schlechter klassifizierte als bei einer zufälligen Klassifikation zu erwarten.

Darüber hinaus war von Interesse, ob sich einige der Variablen zwischen den Gruppen statistisch signifikant unterschieden. Vor allem für die traditionell verwendete Variable *nr_responses* sollte untersucht werden, ob die Gruppe der Patienten mit KKD weniger korrekte Nennungen im Vergleich zur Gruppe der Sprachgesunden produzierte. Alle statistischen Analysen wurden mit der Statistiksoftware PASW Statistics 18 (2009) gerechnet.

Die Ergebnisse zeigten keine signifikanten Gruppenunterschiede für die Baseline-Variable *nr_responses* zwischen Patienten mit KKD und Sprachgesunden gemittelt über alle Wortgenerierungsaufgaben (Patienten mit KKD: $M = 12.68$, $SD = 2.53$; Sprachgesunde: $M = 14.39$, $SD = 3.30$; *t*-Test für unabhängige Stichproben: $t(26) = -1.543$, $p = .135$). Das heißt, Patienten mit KKD und Sprachgesunde produzierten eine vergleichbare *Anzahl korrekter Nennungen*.

Des Weiteren wurden die Leistungen entsprechend der Prozentrangwerte aus dem Regensburger Wortflüssigkeitstest (Aschenbrenner et al., 2000) in beiden Gruppen analysiert. Bei der klinischen Auswertung werden die Testwerte (hier: *Anzahl korrekter Nennungen*) in Prozentränge transferiert, um so eine individuelle Leistung im Vergleich zu einer Referenzgruppe betrachten zu können. Es besteht die Konvention, dass Prozentrangwerte kleiner als 16 unterdurchschnittliche Leistungen anzeigen und Prozentrangwerte größer als 84 überdurchschnittliche Leistungen erkennen lassen (Aschenbrenner et al., 2000; Fisseni, 2004). Nach Ermittlung der Prozentrangwerte (entsprechend der Referenzstichprobe in Aschenbrenner et al., 2000) zur Leistungsbeurteilung, zeigten 3 von 14 Sprachgesunden unterdurchschnittliche Leistungen und 9 von 14 Patienten mit kommunikativ-kognitiven Defiziten durchschnittliche Leistungen in den Wortgenerierungsaufgaben. Die-

se Ergebnisse zeigen, wie schwierig es sein kann, zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunde reliabel zu dissoziieren, wenn traditionelle klinische Analysemethoden angewendet werden. Neue Testprozeduren und die Evaluation zusätzlicher Variablen könnten dabei helfen, eine bessere Dissoziation zu ermöglichen.

Abschließend wurde mit statistischen Analysen überprüft, ob signifikante Gruppenunterschiede zwischen den Patienten mit KKD und den Sprachgesunden für die Variablen aus der Liste der besten zehn bestanden (siehe Tabelle 5.2).

Tabelle 5.2: Mittelwerte und Ergebnisse des *t*-Tests unabhängiger Stichproben für die Variablen mit der besten accuracy (gemittelt über alle Aufgaben)

	KKD	Sprachgesund	<i>t</i> -Test	
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
<i>lowest_phon</i>	3.0 (0.20)	2.93 (0.36)	$t(26) = 0.653$	$p = .520$
<i>shortest_break</i>	1.44 (0.66)	0.75 (0.49)	$t(26) = 3.178$	$p < .01^{**}$
<i>highest_phon</i>	9.95 (1.00)	10.21 (1.39)	$t(26) = - 0.585$	$p = .563$
<i>avg_phon</i>	5.51 (0.37)	5.77 (0.60)	$t(26) = - 1.370$	$p = .182$
<i>first_interval</i>	7.93 (1.49)	9.71 (2.06)	$t(26) = - 2.628$	$p < .05^*$
<i>avg_break</i>	4.71 (1.43)	3.93 (1.02)	$t(26) = 1.662$	$p = .109$
<i>avg_google</i>	196.31 (59.90)	139.26 (61.84)	$t(26) = 2.479$	$p < .05^*$
<i>var_syll</i>	0.94 (0.28)	0.92 (0.28)	$t(26) = 0.208$	$p = .837$
<i>var_specificity</i>	4.99 (2.45)	4.37 (1.33)	$t(26) = 0.825$	$p = .417$
<i>highest_syll</i>	3.96 (0.45)	3.93 (0.50)	$t(26) = 0.198$	$p = .844$

Für die Variablen *shortest_break*, *first_interval* und *avg_google* konnten signifikante Gruppenunterschiede gefunden werden. Die Analysen ergaben, dass die kürzeste Pause in der Gruppe der Sprachgesunden signifikant kürzer war als in der Gruppe der Patienten mit KKD. Die *Anzahl korrekter Nennungen* war im ersten Interval für die Gruppe der Sprachgesunden signifikant höher als für die Gruppe der Patienten mit KKD. Darüber hinaus hatten Patienten mit KKD eine signifikant größere durchschnittliche Anzahl von *Google Hits*, das heißt, sie produzierten im Vergleich zu den Sprachgesunden signifikant mehr hochfrequente Wörter. Die Ergebnisse weisen darauf hin, dass die Variablen *shortest_break*, *first_interval* und *avg_google* geeignet sein könnten, um Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunde voneinander abzugrenzen.

5.3.2 Experiment 1b

Traditionell werden die verschiedenen Wortgenerierungsaufgaben unter Verwendung desselben Maßes ausgewertet. Die verschiedenen Aufgaben untersuchen jedoch zum Teil divergente kognitive Prozesse und sind mit Aktivierung in unterschiedlichen Hirnregionen assoziiert (Basso et al., 1997; Nutter-Upham et al., 2008; Troyer et al., 1998). Es ist somit anzunehmen,

das bestimmte Variablen für die Evaluation von Leistungen in einer Wortgenerierungsaufgabe besser geeignet sein könnten als für die Analyse der Leistungen in einer anderen Aufgabe. Daher wurde die Effizienz der Variablen auch individuell für die vier verschiedenen Wortgenerierungsaufgaben untersucht.

Methoden

Wie in Experiment 1a, wurde auch hier ein 1NN Klassifikator individuell für jede Wortgenerierungsaufgabe und jede Variable trainiert. Anschließend wurde die erzielte Klassifikationsgenauigkeit der verschiedenen Variablen miteinander verglichen; dieses Mal ohne vorher über alle Wortgenerierungsaufgaben zu mitteln.

Ergebnisse & Diskussion

Tabelle 5.3 zeigt eine Liste der besten fünf Variablen für jede Wortgenerierungsaufgabe; das heißt, die Variablen, die die beste *accuracy* erzielten. Erreichten mehrere Variablen dieselbe *accuracy*, sind alle genannt, was bedeutet, dass auch mehr als fünf Variablen gelistet sein können. Zusätzlich ist für jede Wortgenerierungsaufgabe auch die *accuracy* der Baseline-Variablen *nr_responses* angegeben.

Es ist zu erkennen, dass für jede Wortgenerierungsaufgabe andere Variablen die beste Klassifikationsgenauigkeit erreichten. Auch die jeweils erzielten Werte waren in den einzelnen Wortgenerierungsaufgaben unterschiedlich. Mit den semantischen Wortgenerierungsaufgaben konnten im Vergleich zu den beiden lexikalischen Aufgaben bessere Ergebnisse erreicht werden. Die beste *accuracy* in Höhe von 75.0% erzielten die Variablen *avg_break* und *lowest_phon* in der simplen semantischen Aufgabe (SEM_Tiere). Die Baseline-Variablen *nr_responses* war für keine der vier Wortgenerierungsaufgaben unter den als beste gelisteten Variablen. Bei drei von vier Aufgaben (LEX_M, SEM_Tiere und LEX_H/T) zeigte die Baseline-Variablen mit einer *accuracy* kleiner als 50 Prozent sogar schlechtere Leistungen als mit einer zufälligen Klassifikation hätte erreicht werden können. Bei der Aufgabe SEM_Kleidung/Blumen entsprach die Leistung der Variablen exakt der, die auch durch eine Zufallsklassifikation erreicht werden könnte. Es lässt sich festhalten, dass die Variablen *nr_responses* bei keiner der vier Wortgenerierungsaufgaben die Klassifikation faszilierte.

Die Ergebnisse weisen darauf hin, dass die Verwendung unterschiedlicher, das heißt aufgabenspezifischer Variablen bei der Analyse verschiedener Wortgenerierungsaufgaben zu besseren Klassifikationsergebnissen führen kann, als die Verwendung derselben Variablen bei allen Aufgaben. Die maximale Klassifikationsgenauigkeit, die die Variablen *nr_responses* bei der

5 Exploration einer Methodologie zur Evaluation von Wortgenerierungsleistungen mit *machine learning*-Techniken

Tabelle 5.3: Liste der Variablen mit der besten *accuracy* in den vier Wortgenerierungsaufgaben - Experiment 1b

LEX_M		SEM_Tiere	
Variablen	Accuracy	Variablen	Accuracy
<i>var_google</i>	67.86	<i>avg_break</i>	75.0
<i>var_phon</i>	64.29	<i>lowest_phon</i>	75.0
<i>lowest_phon</i>	64.29	<i>nr_switches</i>	67.86
<i>var_frequency</i>	60.71	<i>highest_phon</i>	67.86
<i>longest_break</i>	57.14	<i>first_interval</i>	64.29
<i>avg_frequency</i>	57.14	<i>shortest_break</i>	64.29
<i>lowest_syll</i>	57.14	<i>var_wiki</i>	64.29
<i>first_interval</i>	57.14	<i>avg_google</i>	64.29
<i>var_syll</i>	57.14		
<i>avg_phon</i>	57.14		
<i>var_syll</i>	57.14		
<i>lowest_frequency</i>	57.14		
<i>var_break</i>	57.14		
<i>lowest_specificity</i>	57.14		
<i>nr_responses</i>	39.29	<i>nr_responses</i>	46.43
LEX_H/T		SEM_Kleidung/Blumen	
Variablen	Accuracy	Variablen	Accuracy
<i>highest_phon</i>	64.29	<i>shortest_break</i>	71.43
<i>shortest_break</i>	60.71	<i>avg_break</i>	67.86
<i>lowest_google</i>	60.71	<i>avg_phon</i>	64.29
<i>highest_frequency</i>	60.71	<i>var_syll</i>	64.29
<i>avg_syll</i>	57.14	<i>highest_syll</i>	57.14
<i>avg_phon</i>	57.14	<i>lowest_phon</i>	57.14
<i>lowest_phon</i>	57.14	<i>highest_phon</i>	57.14
<i>highest_specificity</i>	57.14		
<i>nr_responses</i>	21.43	<i>nr_responses</i>	50.0

Betrachtung einer einzelnen Wortgenerierungsaufgabe erzielte, lag bei 50% (für die Aufgabe SEM_Kleidung/Blumen). Die Verwendung anderer Variablen führte zu einer Verbesserung der Klassifikationsgenauigkeit. Dies zeigt, dass andere Variablen als *nr_responses* besser geeignet sein könnten, um zwischen Patienten mit KKD und Sprachgesunden zu dissoziieren. Darüber hinaus scheint es bei Anwendung von *machine learning*-Techniken angemessen, nur einen Teil der Wortgenerierungsaufgaben zu evaluieren. Die Auswertung der Aufgabe SEM_Tiere könnte dabei vielversprechend sein, da zwei Variablen eine *accuracy* von 75% erzielten, was einer Verbesserung von etwas mehr als 10% im Vergleich zu den besten Variablen in den anderen drei Wortgenerierungsaufgaben entsprach.

Wie in Experiment 1a waren auch hier Gruppenunterschiede für die traditionell verwendete Baseline-Variable *nr_responses* von Interesse. Es wurde untersucht, ob die Gruppe der Patienten mit KKD in den einzelnen Wortgenerierungsaufgaben signifikant weniger korrekte Nennungen produ-

zierte als die Kontrollgruppe der Sprachgesunden. Tabelle 5.4 zeigt keine reliablen Gruppenunterschiede zwischen Patienten mit KKD und Sprachgesunden für die Variable *nr_responses* in den einzelnen Aufgaben. Das heißt, beide Gruppen produzierten eine vergleichbare *Anzahl korrekter Nennungen* in allen vier Wortgenerierungsaufgaben.

Tabelle 5.4: *Mittelwerte und Ergebnisse des t-Tests unabhängiger Stichproben für die Baseline-Variable nr_responses in den vier Wortgenerierungsaufgaben*

	KKD	Sprachgesund	<i>t</i> -Test	
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
LEX_M	8.86 (3.28)	9.71 (4.21)	<i>t</i> (26) = - 0.601	<i>p</i> = .553
SEM_Tiere	19.36 (5.87)	22.57 (5.69)	<i>t</i> (26) = - 1.471	<i>p</i> = .153
LEX_H/T	11.64 (3.00)	12.36 (3.52)	<i>t</i> (26) = - 0.577	<i>p</i> = .569
SEM_Kleidung/ Blumen	10.86 (2.88)	12.93 (2.64)	<i>t</i> (26) = - 1.983	<i>p</i> = .058

Für die einzelnen Wortgenerierungsaufgaben sollte darüber hinaus evaluiert werden, wie viele der Versuchsteilnehmer einer jeden Gruppe *unterdurchschnittliche*, *durchschnittliche* oder *überdurchschnittliche* Wortgenerierungsleistungen zeigten. Die Evaluation der Prozenträge (entsprechend der Referenzstichprobe in Aschenbrenner et al., 2000) ergab die folgenden Leistungsbeurteilungen (vgl. Tabelle 5.5):

Tabelle 5.5: *Einschätzung der Wortgenerierungsleistung im Vergleich zur Referenzstichprobe aus dem Regensburg Wortflüssigkeitstest (Aschenbrenner et al., 2000) in den vier Wortgenerierungsaufgaben für Patienten mit KKD und Sprachgesunde; Angegeben sind jeweils die Häufigkeiten unterdurchschnittlicher (ULei), durchschnittlicher (DLei) sowie überdurchschnittlicher Leistung (ÜLei) in den beiden Gruppen*

Aufgabe	Gruppe	ULei	DLei	ÜLei
LEX_M	Sprachgesund	7	6	1
	KKD	6	8	0
SEM_Tiere	Sprachgesund	3	10	1
	KKD	5	9	0
LEX_H/T	Sprachgesund	5	9	0
	KKD	5	8	1
SEM_Kleidung/ Blumen	Sprachgesund	3	10	1
	KKD	6	8	0

Die Leistungsbeurteilungen für die einzelnen Wortgenerierungsaufgaben zeigten, dass es weiterhin eine Herausforderung war, Patienten mit KKD von Sprachgesunden anhand der Leistungen bei der Wortgenerierung zu unterscheiden, wenn ausschließlich traditionelle Maße verwendet wurden. Sogar sprachgesunde Versuchsteilnehmer zeigten in allen vier Wortgenerierungsaufgaben unterdurchschnittliche Leistungen.

Zusätzliche statistische Analysen sollten überprüfen, ob es hinsichtlich der als beste gelisteten Variablen in den einzelnen Aufgaben (vgl. Tabelle 5.3) signifikante Gruppenunterschiede zwischen den Patienten mit KKD und den Sprachgesunden gab. Tabelle 5.6 zeigt die Ergebnisse für die Aufgabe *LEX_M*: signifikante Gruppenunterschiede konnten für keine der Variablen gefunden werden.

Tabelle 5.6: Mittelwerte und Ergebnisse des *t*-Tests unabhängiger Stichproben für die Variablen mit der besten *accuracy* in Aufgabe *LEX_M*

	KKD	Sprachgesund	<i>t</i> -Test	
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
<i>var_google</i>	483219.69 (1388712.62)	118717.57 (233516.75)	$t(26) = 0.968$	$p = .342$
<i>var_phon</i>	6.08 (3.81)	5.83 (5.03)	$t(26) = 0.147$	$p = .884$
<i>lowest_phon</i>	3.07 (0.27)	3.43 (0.65)	$t(17.32) = - 1.911$	$p = .073$
<i>var_frequency</i>	7772.98 (13771.39)	104107.07 (363387.64)	$t(26) = - 0.991$	$p = .331$
<i>longest_break</i>	12.32 (4.50)	15.05 (6.77)	$t(26) = - 1.260$	$p = .219$
<i>avg_frequency</i>	42.53 (40.15)	61.49 (95.79)	$t(26) = - 0.683$	$p = .501$
<i>lowest_syll</i>	1.00 (0.00)	1.14 (0.36)	$t(13) = - 1.472$	$p = .165$
<i>first_interval</i>	5.36 (1.98)	6.14 (2.82)	$t(26) = - 0.852$	$p = .402$
<i>var_syll</i>	1.20 (0.67)	1.06 (0.57)	$t(26) = 0.584$	$p = .564$
<i>avg_phon</i>	5.81 (1.07)	6.41 (1.14)	$t(26) = - 1.427$	$p = .165$
<i>lowest_frequency</i>	0.00 (0.00)	0.14 (0.36)	$t(13) = - 1.472$	$p = .165$
<i>var_break</i>	16.94 (18.11)	26.92 (30.61)	$t(26) = - 1.050$	$p = .303$
<i>lowest_specificity</i>	3.14 (0.66)	2.79 (1.19)	$t(20.38) = 0.982$	$p = .338$

Die Ergebnisse für die Aufgabe *SEM_Tiere* sind in Tabelle 5.7 angegeben. Signifikante Gruppenunterschiede zeigten sich für die Variablen *avg_break*, *lowest_phon*, *nr_switches* und *first_interval*. Die Nennungen der Sprachgesunden wiesen sowohl eine signifikant kürzere durchschnittliche Pausenlänge, als auch eine signifikant geringere kleinste Phonemanzahl als die von Patienten mit KKD auf. Die Gruppe der Sprachgesunden produzierte darüber hinaus eine signifikant größere Anzahl von *Switches* sowie signifikant mehr korrekte Nennungen in den ersten 30 Sekunden als die Gruppe der Patienten mit KKD.

Keine signifikanten Gruppenunterschiede bzgl. der gelisteten Variablen wurden für die lexikalische Wechsel-Aufgabe (*LEX_H/T*; vgl. Tabelle 5.8) gefunden.

Tabelle 5.9 zeigt signifikante Gruppenunterschiede in der semantischen Wechsel-Aufgabe *SEM_Kleidung/Blumen* für die Variablen *shortest_break*, *avg_phon* und *highest_phon*. Die Nennungen der Sprachgesunden wiesen eine signifikant kürzere längste Pause, eine höhere durchschnittliche Anzahl von Phonemen sowie eine höhere größte Anzahl von Phonemen im Vergleich zu denen der Gruppe der Patienten mit KKD auf.

Tabelle 5.7: Mittelwerte und Ergebnisse des *t*-Tests unabhängiger Stichproben für die Variablen mit der besten accuracy in Aufgabe SEM_Tiere

	KKD	Sprachgesund	<i>t</i> -Test	
	<i>M (SD)</i>	<i>M (SD)</i>		
<i>avg_break</i>	2.64 (1.12)	1.83 (0.51)	$t(18.15) = 2.466$	$p < .05^*$
<i>lowest_phon</i>	2.79 (0.43)	2.29 (0.47)	$t(26) = 2.954$	$p < .01^{**}$
<i>nr_switches</i>	7.21 (2.61)	9.57 (3.34)	$t(26) = - 2.080$	$p < .05^*$
<i>highest_phon</i>	9.86 (2.21)	9.07 (1.49)	$t(26) = 1.101$	$p = .281$
<i>first_interval</i>	11.57 (3.61)	15.29 (2.97)	$t(26) = - 2.972$	$p < .01^{**}$
<i>shortest_break</i>	0.28 (0.44)	0.08 (0.16)	$t(16.23) = 1.669$	$p = .114$
<i>var_specificity</i>	5.28 (2.13)	5.17 (2.80)	$t(26) = 0.116$	$p = .909$
<i>avg_google</i>	74.64 (62.96)	67.99 (62.34)	$t(26) = 0.281$	$p = .781$

Tabelle 5.8: Mittelwerte und Ergebnisse des *t*-Tests unabhängiger Stichproben für die Variablen mit der besten accuracy in Aufgabe LEX_H/T

	KKD	Sprachgesund	<i>t</i> -Test	
	<i>M (SD)</i>	<i>M (SD)</i>		
<i>highest_phon</i>	9.79 (1.12)	9.93 (2.62)	$t(17.63) = - 0.188$	$p = .853$
<i>shortest_break</i>	1.67 (1.22)	0.98 (1.01)	$t(26) = 1.639$	$p = .113$
<i>lowest_google</i>	2.83 (5.30)	2.36 (3.81)	$t(26) = 0.269$	$p = .790$
<i>highest_frequency</i>	520.43 (879.97)	148.57 (102.24)	$t(13.35) = 1.571$	$p = .140$
<i>avg_syll</i>	2.00 (0.21)	2.16 (0.49)	$t(17.42) = - 1.163$	$p = .260$
<i>avg_phon</i>	5.30 (0.45)	5.69 (1.18)	$t(16.62) = - 1.165$	$p = .261$
<i>lowest_phon</i>	2.93 (0.27)	3.00 (0.55)	$t(26) = - 0.434$	$p = .668$
<i>highest_specificity</i>	11.07 (3.60)	10.36 (1.69)	$t(26) = 0.671$	$p = .508$

Tabelle 5.9: Mittelwerte und Ergebnisse des *t*-Tests unabhängiger Stichproben für die Variablen mit der besten accuracy in Aufgabe SEM_Kleidung/Blumen

	KKD	Sprachgesund	<i>t</i> -Test	
	<i>M (SD)</i>	<i>M (SD)</i>		
<i>shortest_break</i>	1.89 (1.42)	0.78 (0.68)	$t(26) = 2.635$	$p < .05^*$
<i>avg_break</i>	5.36 (2.94)	3.81 (1.05)	$t(26) = 1.864$	$p = .074$
<i>avg_phon</i>	5.47 (0.47)	5.88 (0.42)	$t(26) = - 2.456$	$p < .05^*$
<i>var_syll</i>	0.88 (0.33)	0.99 (0.27)	$t(26) = - 1.026$	$p = .314$
<i>highest_syll</i>	4.00 (0.68)	4.07 (0.27)	$t(26) = - 0.366$	$p = .717$
<i>lowest_phon</i>	3.21 (0.43)	3.00 (0.68)	$t(26) = 1.000$	$p = .327$
<i>highest_phon</i>	9.79 (1.81)	11.50 (1.51)	$t(26) = - 2.728$	$p < .05^*$

Zusammengenommen lässt sich festhalten: Es fanden sich keine signifikanten Gruppenunterschiede bezüglich der Variablen mit der besten Klassifikationsgenauigkeit für die simple (LEX_M) und die komplexe (LEX_H/T) lexikalischen Wortgenerierungsaufgaben. Für die beiden semantischen Aufgaben SEM_Tiere und SEM_Kleidung/Blumen konnten dagegen für mehrere der als beste gelisteten Variablen signifikante Gruppenunterschiede ge-

funden werden. Darüber hinaus erzielte die Evaluation der Variablen mit *machine-learning*-Techniken (vgl. Tabelle 5.3) bessere Klassifikationsergebnisse für die beiden semantischen Wortgenerierungsaufgaben im Vergleich zu den lexikalischen. Die Ergebnisse deuten darauf hin, dass semantische Wortgenerierungsaufgaben besser geeignet sein könnten, um zwischen Patienten mit KKD und Sprachgesunden zu dissoziieren.

5.4 Experiment 2

Die Ergebnisse aus den beiden vorherigen Experimenten weisen darauf hin, dass andere Variablen als die *Anzahl korrekter Nennungen* besser geeignet sein könnten, um Patienten mit KKD von Sprachgesunden zu dissoziieren. Zu diesen zählte beispielsweise in Experiment 1a die zeitbasierte Variable *shortest_break* (siehe z.B. Tabelle 5.1 und Tabelle 5.2). Im Folgenden wird untersucht, ob die Verwendung von Variablenkombinationen die Klassifikation weiter verbessert.

5.4.1 Experiment 2a

Ähnlich der Evaluation einzelner Variablen, wurden zunächst Variablenkombinationen ermittelt, die für die Evaluation von allen Wortgenerierungsaufgaben am besten geeignet schienen. Anschließend wurde die Leistung der Variablenkombinationen, die gemittelt über alle Wortgenerierungsaufgaben die beste Performanz zeigten, mit der Leistung der einzelnen Variablen, die die beste Leistung gemittelt über alle Aufgaben zeigten, verglichen.

Methoden

Für alle möglichen Variablenpaare und Variablentripel wurde ein 1NN Klassifikator trainiert. Die Berechnung von Ergebnissen für alle Paare und Tripel garantierte, dass das beste Variablen-Set bestehend aus je zwei bzw. drei Variablenkombinationen gefunden werden konnte. Aufgrund der erforderlichen Rechenleistung und da die Verwendung zusätzlicher Variablen die Performanz eines Klassifikators verringern kann, wenn die Anzahl von Datensätzen im Vergleich zur Anzahl der Variablen relativ gering ist (Jain et al., 2000), wurden nur Kombinationen aus wenigen Variablen verwendet. Die Ergebnisse für die jeweiligen Variablenkombinationen wurden anschließend über alle Wortgenerierungsaufgaben gemittelt und entsprechend ihrer *accuracy* gelistet.

Ergebnisse & Diskussion

Die Berechnungen ergaben insgesamt bessere Ergebnisse für die Variablentripel im Vergleich zu den Variablenpaaren. Daher wird sich die Darstellung

der Ergebnisse im Folgenden auf die Variablenkombinationen fokussieren.³⁵ Eine Liste der besten zehn Variablenkombinationen, gemittelt über alle Wortgenerierungsaufgaben, ist in Tabelle 5.10 abgebildet.

Tabelle 5.10: *Liste der Variablenkombinationen mit der besten accuracy gemittelt über alle Aufgaben - Experiment 2a*

Variablen	Accuracy
<i>avg_syll, highest_phon, shortest_break</i>	68.75
<i>avg_google, highest_frequency, shortest_break</i>	68.75
<i>avg_phon, highest_phon, shortest_break</i>	68.75
<i>avg_syll, avg_specificity, highest_phon</i>	67.86
<i>avg_syll, highest_phon, lowest_syll</i>	67.86
<i>highest_phon, lowest_phon, var_frequency</i>	67.86
<i>avg_phon, highest_phon, var_frequency</i>	67.86
<i>avg_phon, highest_specificity, shortest_break</i>	67.85
<i>avg_specificity, highest_phon, lowest_phon</i>	66.97
<i>avg_phon, highest_phon, highest_specificity</i>	66.96

Alle Variablenkombinationen aus der Liste der besten zehn erzielten eine bessere *accuracy* als die, die mit den besten Einzelvariablen (vgl. Tabelle 5.1) erreicht werden konnten. Verglichen mit der besten Einzelvariable (*accuracy* = 63.39%), erzielten die besten drei Variablenkombinationen eine *accuracy* in Höhe von 68.75%, was einer Verbesserung von etwa 5% entspricht. Die Baseline-Variablenkombination *nr_responses* ist in keinem gelisteten Variablenkombinationen enthalten.

Die Ergebnisse legen nahe, dass bei der Evaluation von Wortgenerierungsleistungen Variablenkombinationen im Vergleich zu Einzelvariablen besser geeignet sind, um zwischen Patienten mit KKD und Sprachgesunden zu dissoziieren. Darüber hinaus zeigt ein Vergleich von Tabelle 5.1 und Tabelle 5.10, dass einige der als beste gelisteten Einzelvariablen (z.B. *highest_phon*, *shortest_break*, *avg_phon*) auch in der Liste der besten Variablenkombinationen vertreten sind. Insbesondere solche Variablen könnten gut geeignet sein, um zwischen Patienten mit KKD und Sprachgesunden zu dissoziieren.

5.4.2 Experiment 2b

Experiment 1b ergab, dass verschiedene Einzelvariablen unterschiedlich gut in den jeweiligen Wortgenerierungsaufgaben zwischen Patienten mit KKD und Sprachgesunden dissoziierten. Im Folgenden soll daher überprüft werden, ob dies auch auf die Analyse mit Variablenkombinationen zutrifft.

³⁵Wird im Folgenden von Variablenkombinationen gesprochen, sind stets Variablenkombinationen, also Kombinationen aus drei Variablen gemeint.

Methoden

Individuell für jedes Variablen-Tripel und jede Wortgenerierungsaufgabe wurde ein 1NN Klassifikator trainiert. Die Variablen-Tripel wurden anschließend entsprechend der ermittelten *accuracy* gelistet, jedoch dieses Mal ohne über die Aufgaben zu mitteln.

Die vorausgehenden Berechnungen wiesen darauf hin, dass Variablen-Tripel anstelle von Einzelvariablen sowie eine individuelle Evaluation der Aufgaben anstelle einer Gesamtauswertung, die Klassifikationsgenauigkeit verbessert. Um einen Einblick zu gewinnen, wie effizient die eingeführten Variablen für die Evaluation von Wortgenerierungsaufgaben waren (vor allem in Bezug auf die Anwendbarkeit bei der klinischen Diagnosestellung), wurden zusätzlich zur Klassifikationsgenauigkeit zwei weitere Evaluationsmaße ermittelt. Diese wurden für jede der beiden Gruppen berechnet, das heißt für die Gruppe der Patienten mit KKD und die Gruppe der Sprachgesunden:

- *Precision* (auch: *positiver Vorhersagewert*): Dieses Maß ist definiert als Anzahl korrekt klassifizierter positiver Instanzen geteilt durch die Anzahl aller Instanzen, die positiv klassifiziert wurden (das heißt: richtig Positive / (richtig Positive + falsch Positive). *Precision* ist ein wichtiges Maß in der klinischen Analyse, da es die Wahrscheinlichkeit angibt, mit der ein positives Testergebnis die korrekte Diagnose widerspiegelt (Bortz & Lienert, 2008). Es gibt also die Wahrscheinlichkeit wieder, mit der eine Person deren Testergebnis ein kommunikativ-kognitives Defizit anzeigt, auch wirklich ein Patient mit einer kommunikativ-kognitiven Beeinträchtigung ist.
- *Recall* (auch: *Richtig-Positiv-Rate*): Dieses Maß ist definiert als korrekt klassifizierte, positive Instanzen geteilt durch die Gesamtzahl positiver (das heißt: richtig Positive / (richtig Positive + falsch Negative). Dieses Maß wird häufig auch als *Sensitivität* (Bortz & Lienert, 2008) bezeichnet und zeigt die Wahrscheinlichkeit an, mit der eine positive Diagnose mit einem Test bestätigt werden kann. Es spiegelt also z.B. die Wahrscheinlichkeit wider, mit der eine Person mit kommunikativ-kognitivem Defizit auch ein Testergebnis erhält, das eine kommunikativ-kognitive Beeinträchtigung bestätigt.

Ergebnisse & Diskussion

Im Folgenden findet sich eine Übersicht der besten Variablen-Tripel für die einzelnen Wortgenerierungsaufgaben (vgl. Tabelle 5.11).

Ähnlich dem Training auf Basis von Einzelvariablen unterschieden sich auch die Variablen-Tripel, die die besten Klassifikationsergebnisse erzielten,

Tabelle 5.11: Liste der besten Variablenkombinationen mit Angaben zu Accuracy, Precision (Prec) und Recall (Rec) für die einzelnen Wortgenerierungsaufgaben - Experiment 2b

	LEX_M		SEM_Tiere		SEM_Kleidung/ Blumen	
Variablen	<i>highest_frequency</i> <i>lowest_specificity</i> <i>avg_syll</i>		<i>highest_phon</i> <i>lowest_phon</i> <i>avg_specificity</i>		<i>highest_phon</i> <i>avg_syll</i> <i>shortest_break</i>	
Accuracy	85.71		89.29		89.29	
	Prec	Rec	Prec	Rec	Prec	Rec
KKD	0.86	0.86	0.92	0.86	0.87	0.93
Sprachgesund	0.86	0.96	0.87	0.93	0.92	0.86
	LEX_H/T					
Variablen	<i>highest_frequency</i> <i>shortest_break</i> <i>avg_google</i>		<i>highest_phon</i> <i>second_interval</i> <i>avg_syll</i>			
Accuracy	89.29		89.29			
	Prec	Rec	Prec	Rec		
KKD	0.87	0.93	1.0	0.79		
Sprachgesund	0.92	0.86	0.82	1.0		

in den verschiedenen Wortgenerierungsaufgaben. Drei von vier Wortgenerierungsaufgaben erzielten mit unterschiedlichen Variablenkombinationen die besten und gleichzeitig auch identische Ergebnisse. Die besten Variablen-triplet der Aufgaben SEM_Tiere, LEX_H/T und SEM_Kleidung/Blumen erzielten eine *accuracy* in Höhe von 89.29%. Mit der Aufgabe LEX_M wurde eine leicht geringere Klassifikationsgenauigkeit in Höhe von 85.71% erzielt. Diese Ergebnisse stellten eine deutliche Verbesserung bezüglich der Klassifikationsgenauigkeit im Vergleich zum Training auf Basis von Einzelvariablen dar. Im Vergleich zur maximalen *accuracy* in Höhe von 50% für die Baseline-Variable *nr_responses* bei der Evaluation der einzelnen Wortgenerierungsaufgaben (vgl. Tabelle 5.3), zeigte sich hier eine Verbesserung des Klassifikationsergebnisses in Höhe von fast 40%. Es war jedoch auch zu erkennen, dass, obwohl in drei Aufgaben ein identisches Klassifikationsergebnis (*accuracy* = 89.29%) erzielt werden konnte, sich die Werte für *Precision* und *Recall* in diesen Aufgaben dennoch unterschieden.

Bis jetzt weisen die Ergebnisse darauf hin, dass die Evaluation einzelner Wortgenerierungsaufgaben bei Verwendung von Variablenkombinationen zu vielversprechenden Ergebnissen führt. Die Werte für *Precision* und *Recall* unterschieden sich jedoch in den einzelnen Wortgenerierungsaufgaben zwischen den beiden Gruppen. Daher scheint es aussichtsreich, eine spezifische Wortgenerierungsaufgabe für die Evaluation auszuwählen, insbesondere wenn Variablen zusätzlich zur *accuracy* wichtig sind. Wenn das Ziel z.B. darin besteht, eine Methode zu entwickeln, die die Diagnose eines KKD sicher

bestätigen kann, scheint die Evaluation der Aufgabe *LEX_H/T* mit dem Variablen-Tripel *highest_phon*, *second_interval*, *avg_syll* geeignet, da mit dieser Kombination für die Gruppe der Patienten mit KKD eine *Precision* in Höhe von 100% erzielt werden konnte. Das bedeutet, dass jeder Patient, der als Patient mit KKD klassifiziert wurde, auch tatsächlich ein Patient mit kommunikativ-kognitiver Beeinträchtigung ist. Demgegenüber zeigt der *Recall*-Wert (79%) an, dass nicht alle Patienten mit KKD identifiziert werden konnten, sondern einige Patienten mit KKD als Sprachgesunde fehlklassifiziert wurden. Die Entwicklung einer solchen Methode erscheint trotzdem für die klinische Praxis dienlich, da sie für einen Großteil der Patienten eine reliable Identifizierung einer kommunikativ-kognitiven Beeinträchtigung möglich macht. Weitere diagnostische Prozeduren müssten dann lediglich Aufschluss über die als sprachgesund fehlklassifizierten Patienten liefern.

Auffällig war, dass einige Variablen, die Teil der Variablen-Tripel mit den besten Klassifikationsergebnissen in den Aufgaben waren (vgl. Tabelle 5.11), auch unter den Besten bei der Evaluation individueller Variablen in den einzelnen Wortgenerierungsaufgaben vertreten waren (vgl. Tabelle 5.3). Einige der Variablen fanden sich bei der Evaluation der meisten Wortgenerierungsaufgaben wieder. So zum Beispiel die Variable *highest_phon* in den Aufgaben SEM_Tiere, LEX_H/T und SEM_Kleidung/Blumen und *shortest_break* in den Aufgaben LEX_H/T und SEM_Kleidung/Blumen. Für die Aufgabe LEX_M tauchte eine andere Variable erneut auf, nämlich *lowest_specificity*. Diesbezüglich unterschied sich die Aufgabe LEX_M von den anderen Wortgenerierungsaufgaben. Schaut man sich die Ergebnisse der statistischen Analyse in Experiment 1b an (vgl. Tabelle 5.6), zeigt sich, dass sich die beiden Gruppen nicht bezüglich der Variable *lowest_specificity* unterschieden. Eine plausible Erklärung warum in der Aufgabe LEX_M andere Variablen auftauchten als in den anderen Aufgaben könnte eventuell auch auf methodische Aspekte zurückzuführen sein. Die Aufgabe LEX_M war stets die erste nach der Übungsaufgabe. Eine Erklärung könnte daher sein, dass Patienten bei dieser Aufgabe noch im Übungsmodus waren und sich die Leistungen bzw. die identifizierten Variablen daher von denen anderer Aufgaben unterschieden.

5.5 Experiment 3

Die beiden vorausgehenden Experimente legen nahe, dass für die jeweiligen Wortgenerierungsaufgaben unterschiedliche Einzelvariablen bzw. Variablenkombinationen die Klassifikationsgenauigkeit im Vergleich zur Baseline-Variablen *nr_responses* verbessern. Bei der Verwendung von Variablen-Tripel konnte für drei von vier Wortgenerierungsaufgaben dieselbe *accuracy* erzielt werden. Im Folgenden wird nun untersucht, ob ein Zusammenführen der

Ergebnisse von Klassifikatoren die auf den einzelnen Wortgenerierungsaufgaben trainiert wurden (im Sinne einer Meta-Klassifikation), eine zusätzliche Verbesserung der Klassifikationsgenauigkeit bewirken kann.

Methode

In Experiment 2b wurde die beste *accuracy* für die Aufgaben SEM_Tiere, LEX_H/T und SEM_Kleidung/Blumen erzielt. Diese drei Aufgaben wurden im folgenden Experiment genutzt, um individuelle Klassifikatoren zu bilden, deren Entscheidungen anschließend verwendet wurden, um eine Gesamtentscheidung im Sinne einer *joint decision*, zu erreichen. Für die Kombination der einzelnen Klassifikatoren wurden zwei verschiedene Strategien exploriert – eine relativ simple und eine komplexere: (i) *Majority voting* und (ii) *Support Vector Machines (SVM)*, ein linearer Kernel mit Platt's *minimal optimization algorithm for training* Platt, 1998). Für jeden Durchgang wurde mit den Trainingsdaten ein 1NN Klassifikator separat für jede Wortgenerierungsaufgabe und den aufgabenspezifischen Variablenkombinationen (vgl. Abschnitt 5.11) gebildet. Während der anschließenden Klassifikation wurde jede Testinstanz von allen Klassifikatoren klassifiziert, das heißt, die Daten jeder Instanz in jeder Wortgenerierungsaufgabe wurden mit dem korrespondierenden Klassifikator der entsprechenden Wortgenerierungsaufgabe klassifiziert. Die Entscheidungen der einzelnen Klassifikatoren wurden anschließend durch *Majority Voting* oder eine *SVM* miteinander kombiniert (= Meta-Klassifikation). Im Fall des *Majority Voting*, führte das Klassifikationsergebnis, das von der Mehrheit der drei Klassifikatoren getroffen wurde, zur Gesamtentscheidung. Abbildung 5.2 veranschaulicht diesen Ablauf.

Im Fall von *SVM* wurden die Entscheidungen der einzelnen Klassifikatoren in eine *SVM* eingespeist, welche dann die Gesamtentscheidung traf. Für jeden Durchgang wurde eine *SVM* auf Basis der Trainingsdaten gebildet. Das heißt im Einzelnen: für die verschiedenen Aufgaben wurden 1NN Klassifikatoren gebildet, bevor anschließend alle Trainingsinstanzen von ihnen klassifiziert und die Ergebnisse verwendet wurden, eine *SVM* zu trainieren. Da für die Aufgabe LEX_H/T zwei Variablen-tripel eine identische *accuracy* erzielten (vgl. Tabelle 5.11), wurden die Berechnungen für beide Kombinationen durchgeführt.

Auch für das beste Variablen-tripel in Aufgabe LEX_M konnte in Experiment 2b ein sehr gutes Klassifikationsergebnis (*accuracy* = 85.71%) erzielt werden. Daher wurde zusätzlich die Performanz der Klassifikatoren bei Verwendung aller Wortgenerierungsaufgaben exploriert. Die Klassifikatoren aller vier Wortgenerierungsaufgaben wurden anhand derselben Strategien miteinander kombiniert wie bei der Verwendung von drei Wortgenerierungsaufgaben.

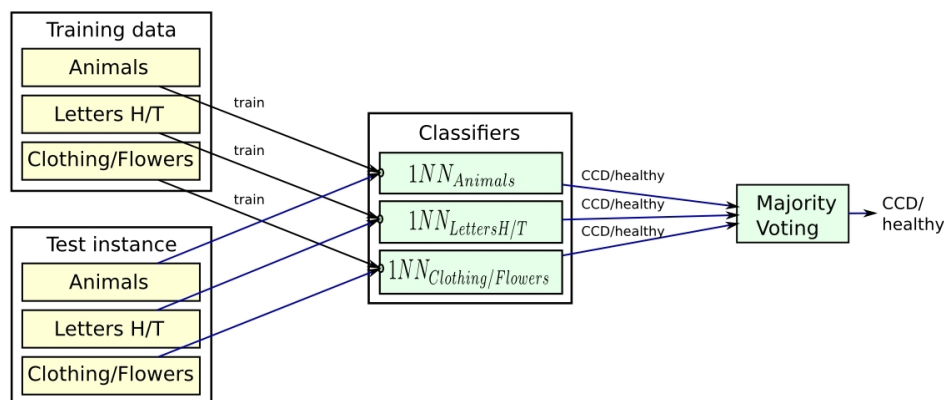


Abbildung 5.2: Meta-Klassifikation mit *Majority Voting*: SEM_Tiere (*Animals*), LEX_H/T (*Letters H/T*), SEM_Kleidung/Blumen (*Clothing/Flowers*), Trainingsdatensatz (*Training Data*), Testinstanzen/ -beispiele (*Test instances*), Training (*train*), Klassifikatoren (*Classifiers*), Patienten mit KKD (*CCD*), Sprachgesund (*healthy*); aus Thiele et al. (under revision, S. 20)

Ergebnisse & Diskussion

Die Ergebnisse der Meta-Klassifikation sind in Tabelle 5.12 abgebildet. Die Bildung eines Meta-Klassifikators auf Basis von drei Wortgenerierungsaufgaben konnte eine weitere Verbesserung der *accuracy* im Vergleich zur Verwendung individueller Klassifikatoren für spezifische Wortgenerierungsaufgaben erzielen. Während die besten Variablenkombinationen in Experiment 2b eine *accuracy* in Höhe von 89.29% erzielten, konnten mit einer Kombination der Klassifikatoren auf Basis von drei Wortgenerierungsaufgaben konsistent *accuracy*-Werte in Höhe von mehr als 90% erzielt werden. Die prozentuale Höhe der *accuracy* beruhte dabei vor allem auf dem Variablenkombination bei der Aufgabe LEX_H/T. Unabhängig vom verwendeten Meta-Klassifikator wurde mit den Variablen aus Kombination 1 eine *accuracy* in Höhe von 96.43% erzielt, eine Verbesserung von etwa 7% im Vergleich zur Klassifikation in Experiment 2b. Mit den Variablen aus Kombination 2 wurde eine Klassifikationsgenauigkeit in Höhe von 92.86% erreicht, was eine Verbesserung von 3.5% im Vergleich zu Experiment 2b darstellt. Das Zusammenführen von Klassifikatoren bzw. deren Entscheidungen aus allen vier Wortgenerierungsaufgaben verbesserte die *accuracy* nicht weiter. Bei der Verwendung von *Majority Voting* war die erzielte *accuracy* bei drei von vier Kombinationen höher als bei der Evaluation einzelner Wortgenerierungsaufgaben in Experiment 2b. Bei Verwendung von SVM war die *accuracy* in allen Fällen höher.

Die Ergebnisse weisen darauf hin, dass ein Trainieren von Klassifikatoren mit individuellen Variablenkombinationen für verschiedene Wortgenerierungsaufgaben und ein anschließendes Zusammenführen im Sinne einer

Tabelle 5.12: *Ergebnisse der Meta-Klassifikation in Experiment 3 - Bei der Verwendung von vier Aufgaben und Majority Voting sind zwei Werte angegeben, da es zwei Stimmen für jede Gruppe gab und somit zwei mögliche Entscheidungen. In der Aufgabe LEX_H/T erzielten zwei Variablenkombinationen die besten Ergebnisse, daher wurden beide untersucht. Kombination 1 beinhaltet die Variablen highest_frequency, shortest_break, avg_google, Kombination 2 die Variablen highest_phon, second_interval, avg_syll.*

Setting	Meta-Klassifikator	Accuracy
Drei Aufgaben, Kombination 1	Majority Voting	96.43
Drei Aufgaben, Kombination 1	SVM	96.43
Drei Aufgaben, Kombination 2	Majority Voting	92.86
Drei Aufgaben, Kombination 2	SVM	92.86
Vier Aufgaben, Kombination 1	Majority Voting	(92.86, 92.86)
Vier Aufgaben, Kombination 1	SVM	92.86
Vier Aufgaben, Kombination 2	Majority Voting	(85.71, 92.86)
Vier Aufgaben, Kombination 2	SVM	92.86

Meta-Klassifikation die beste der explorierten Strategien zu sein scheint, um zwischen Patienten mit KKD und Sprachgesunden reliabel zu dissoziieren. Im Vergleich zu Experiment 1, bei dem die Baseline-Variable *nr_responses* eine *accuracy* in Höhe von 39.29% (gemittelt über alle Aufgaben) und in Höhe von max. 50% (wenn individuell für die Aufgaben berechnet) erzielte, konnte mit der Anwendung eines Meta-Klassifikators in diesem Experiment eine substantielle Verbesserung der Klassifikationsgenauigkeit in Höhe von bis zu 57% bzw. 46% erreicht werden.

Zusammengenommen scheint die Verwendung eines Teils der Wortgenerierungsaufgaben, für die die beste *accuracy* beim Trainieren auf Basis von Variablenkombinationen erzielt wurde, für die Evaluation von Wortgenerierungsleistungen angemessen zu sein. Dadurch konnten, im Vergleich zur Verwendung aller Aufgaben, geringfügig bessere Ergebnisse erreicht werden (vgl. Tabelle 5.12). Insgesamt verringert die Anwendung eines Meta-Klassifikators den erforderlichen Aufwand bei der Evaluation von Wortgenerierungsleistungen, da weniger Wortgenerierungsdaten bezüglich spezifischer Variablen annotiert und für die Variablenauswahl vorbereitet werden müssen. Die Auswertung aufgabenspezifischer Variablen-tripel in der simplen semantischen und den beiden Wechsel-Aufgaben scheinen für eine reliable Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden mit *machine learning*-Techniken nach den Ergebnissen dieser Studie ausreichend.

Obwohl *SVM* und *Majority Voting* ähnliche Ergebnisse erzielt haben, dürfte die *SVM* die bessere Wahl für einen Meta-Klassifikator darstellen: *SVM* erzielte konstant bessere Klassifikationsergebnisse im Vergleich zur Evaluation einzelner Wortgenerierungsaufgaben (Experiment 2b), während

Majority Voting ein Ergebnis mit einer geringeren *accuracy* im Vergleich zur Evaluation in Experiment 2b lieferte. Die Wahl des Meta-Klassifikators sollte darüber hinaus auch von der Anzahl der Aufgaben, die miteinander kombiniert werden sollen, abhängig gemacht werden. *Majority Voting* zeigte nur dann schlechtere Leistungen als *SVM*, wenn eine gerade Anzahl von Aufgaben evaluiert werden sollte, das heißt, wenn die Möglichkeit bestand, dass keine tatsächliche Mehrheit gefunden werden kann (z.B. wenn es eine gleiche Anzahl von Klassifikationen als Patient mit KKD und Sprachgesunder gibt). Besonders im Fall einer gerade Anzahl von Aufgaben scheint die Anwendung einer *SVM* – oder einer ähnlich komplexen Strategie – zu bevorzugen zu sein, da sie auch bei einem ausgeglichenen Voting für Patienten mit KKD und Sprachgesunde lernen kann, ob eine Zuordnung als *Patient mit KKD* oder als *Sprachgesund* erfolgen sollte.

5.6 Diskussion

In dieser Studie wurden drei Experimente präsentiert, die verschiedene Ansätze einer möglichen Evaluationsmethode zur Diagnose kommunikativ-kognitiver Beeinträchtigungen auf der Basis von Wortgenerierungsdaten mittels *machine learning* Techniken explorierten.

Es wurden drei Experimente, mit Daten aus vier verschiedenen Wortgenerierungsaufgaben als Referenzdatenset, durchgeführt. Neben der traditionell bei der klinischen Evaluation verwendeten Variable *Anzahl korrekter Nennungen* wurden weitere zeitliche, sublexikalische, lexikalische und semantische Variablen bei der Analyse berücksichtigt. Im ersten Experiment wurde untersucht, wie effizient Wortgenerierungsleistungen bei Berücksichtigung verschiedener Einzelvariablen evaluiert werden können. Dies sowohl, wenn dieselbe Variable für die Evaluation aller Aufgaben verwendet wird, als auch wenn unterschiedliche Variablen bei der Evaluation der verschiedenen Wortgenerierungsaufgaben angewendet werden. Darüber hinaus wurden Leistungen der zusätzlichen Variablen mit der Leistung des traditionellen Evaluationsmaßes *Anzahl korrekter Nennungen* verglichen. Im zweiten Experiment wurde evaluiert, ob die Verwendung von Variablenkombinationen die Klassifikationsgenauigkeit im Vergleich zur Verwendung einzelner Variablen verbessert. Dies wurde bei Anwendung derselben Variablenkombinationen zur Evaluation aller Aufgaben sowie bei Verwendung aufgabenspezifischer Variablenkombinationen für die Analyse individueller Wortgenerierungsaufgaben untersucht. Experiment drei ging der Frage nach, ob ein Trainieren von Klassifikatoren auf einzelnen Aufgaben und eine anschließende Kombination ihrer Entscheidungen (Meta-Klassifikation) die Klassifikation weiter fazilitiert.

Die beschriebenen Ergebnisse liefern einige Implikationen für die Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden auf Basis der Evaluation von Wortgenerierungsleistungen: Sie weisen darauf hin, dass (i) das Maß, das traditionell in der klinischen Praxis zur Evaluation angewendet wird, das heißt die *Anzahl korrekter Nennungen*, nicht hinreichend gut geeignet zu sein scheint, um kommunikativ-kognitive Defizite verlässlich zu diagnostizieren. Die Ergebnisse zeigen auch, dass (ii) durch die Anwendung aufgabenspezifischer Variablen oder Variablentripel bei der Evaluation von Leistungen in den verschiedenen Wortgenerierungsaufgaben bessere Klassifikationsergebnisse erzielt werden können als bei Anwendung derselben Variablen oder Variablentripel gemittelt über alle Aufgaben. Weiterhin zeigte sich, dass (iii) Variablenkombinationen besser zur Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden geeignet zu sein scheinen als einzelne Variablen und dass (iv) ein Zusammenführen individueller Klassifikationsergebnisse zu einer Gesamtentscheidung (Meta-Klassifikation) die Klassifikationsgenauigkeit weiter verbessert.

Die bedeutendste Einschränkung der vorliegenden Studie liegt in der relativ geringen Größe der untersuchten Stichproben. Aufgrund dieser vergleichsweise geringen Datenmenge sind die Ergebnisse vorerst als explorativ anzusehen. Sobald ein ausreichend großes Datenset zur Verfügung steht, könnte die im Rahmen dieser Arbeit vorgestellte Methodologie jedoch angewendet werden, um eine Evaluationsmethode zu entwickeln, die in der klinischen Praxis anwendbar wäre. Im Folgenden werden nun sowohl Aspekte der Anwendbarkeit, als auch mögliche Erweiterungen zur Verbesserung der Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden im Detail diskutiert.

In den beschriebenen Experimenten konnten verschiedene Variablen identifiziert werden, die für die Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden gut geeignet zu sein scheinen. Insbesondere die Variablen *highest_phon* und *shortest_break* waren häufig entsprechend ihrer Klassifikationsgenauigkeit als beste Variablen gelistet, vor allem in den Aufgaben SEM_Tiere und LEX_H/T sowie in der Aufgabe SEM_Kleidung/Blumen. Diese Variablen könnten für die Klassifikation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden besonders relevant sein und sollten daher auch in zukünftigen Untersuchungen berücksichtigt werden. Aufgrund des relativ kleinen Datensets sollten die Variablen und Variablenkombinationen, die in dieser Studie als geeignete Kandidaten für die Diagnose kommunikativ-kognitiver Defizite identifiziert werden konnten, in zukünftigen Experimenten mit größeren Datensets jedoch validiert werden, bevor spezifischere Aussagen gemacht werden können. Darüber hinaus könnte auch diskutiert werden, zukünftig eine

andere Strategie bei der Variablenselektion zu verwenden. Denkbar wäre hier beispielsweise die Auswahl der Variablen in einer Kreuzvalidierungsschleife (vgl. Maier, Haderlein, Eysholdt, Rosanowski, Batliner, Schuster, & Nöth, 2009).

Darüber hinaus weisen die Ergebnisse darauf hin, dass semantische Wortgenerierungsaufgaben besser zur Dissoziation zwischen Sprachgesunden und Patienten mit kommunikativ-kognitiven Defiziten geeignet sein könnten als lexikalische Aufgaben. Dies stellt ein eher unerwartetes Ergebnis dar, da lexikalische Aufgaben im Vergleich zu semantischen Wortgenerierungsaufgaben als schwieriger beschrieben werden (Azambuja et al., 2007; Basso et al., 1997; Rosser & Hodges, 1994). Lexikalische Wortgenerierungsaufgaben gelten unter anderem deshalb als schwieriger, da die dahinter stehende (phonologische/lexikalische) Suchstrategie nicht als die Strategie angenommen werden kann, die auch in der alltäglichen Kommunikation verwendet wird. Hier wird eher von einer semantischen Strategie ausgegangen (Basso et al., 1997). Basso und Kollegen (1997) beschreiben Folgendes:

”[A] semantic strategy is coherent with the organizational structure of the semantic system and corresponds to the normal way in which we recover words whereas a phonological strategy is quite unusual and it is not used for actual communication” (Basso et al., 1997, S. 544).

Auch Kavé und Kollegen (2011) betonen die Wichtigkeit von semantischen Wortgenerierungsaufgaben für die Evaluation von Wortgenerierungsleistungen. Die Autoren untersuchten die Wortgenerierungsfähigkeit bei einer klinischen Population für die häufig auch von kommunikativ-kognitiven Defiziten berichtet wird: Patienten mit Schädelhirntrauma. Kavé et al. (2011) beschrieben größere Defizite in semantischen im Vergleich zu lexikalischen Aufgaben und nahmen an, dass semantische Aufgaben sensitiver für das Erfassen von Defiziten in den Exekutivfunktionen bei Patienten mit Schädelhirntraumata sein könnten. Zukünftige Studien mit größeren Stichproben sind erforderlich um im Detail aufzudecken, wie Leistungen in semantischen und lexikalischen Aufgaben zur Diagnose kommunikativ-kognitiver Defizite beitragen können.

Es zeigte sich zudem, dass sich die Klassifikationsgenauigkeit erhöht, wenn die jeweiligen Wortgenerierungsaufgaben individuell analysiert werden. Dieses Resultat ist sehr plausibel, da die verschiedenen Aufgaben zum Teil divergente kognitive Prozesse evaluieren und mit Aktivierung in unterschiedlichen Hirnregionen assoziiert sind (Basso et al., 1997; Nutter-Upham et al., 2008; Troyer et al., 1998). Auch andere Autoren betonen die Wichtigkeit der Evaluation von Wortgenerierungsleistungen in verschiedenen Aufgaben (Aschenbrenner et al., 2000; Nutter-Upham et al., 2008). So stellten

etwa Nutter-Upham und Kollegen (2008) die Relevanz der Berücksichtigung unterschiedlicher, das heißt simpler sowie komplexer lexikalischer und semantischer Wortgenerierungsaufgaben bei der Evaluation von Wortgenerierungsleistungen zur Diagnose präklinischer Demenzsymptomatiken heraus. Zusammengenommen lässt sich festhalten, dass die individuelle Betrachtung von Leistungen in verschiedenen Wortgenerierungsaufgaben wichtige Erkenntnisse für Identifizierung spezifischer Störungsbilder zu liefern scheint.

Insgesamt liefern die Ergebnisse der vorliegenden Studie Hinweise darauf, wie eine Methode zur Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden auf Basis einer Evaluation von Wortgenerierungsleistungen konstruiert sein könnte. Die Exploration verschiedener Ansätze in den Experimenten ergab, dass es Erfolg versprechend scheint, Klassifikatoren auf aufgabenspezifischen Variablenkombinationen individuell für die einzelnen Wortgenerierungsaufgaben zu trainieren und ihre Ergebnisse anschließend zu einer Gesamtentscheidung zusammen zu führen. Abbildung 5.3 illustriert den Ablauf eines solchen Vorgehens.

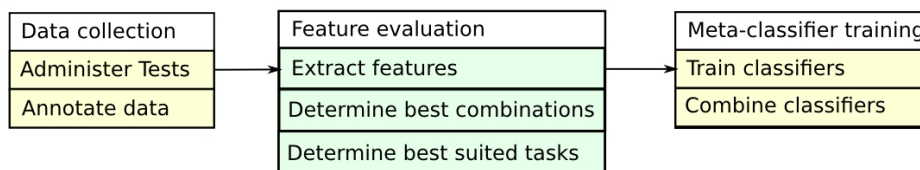


Abbildung 5.3: Vorschlag eines möglichen methodischen Vorgehens: Datenerhebung (*Data Collection*), Erhebung der Wortgenerierungsdaten (*Administer Tests*), Annotation der Daten (*Annotate data*), Evaluation der Variablen (*Feature evaluation*), Extraktion relevanter Variablen aus den Daten (*Extract features*), Bestimmung der besten Variablenkombinationen (*Determine best combinations*), Bestimmung der am besten geeigneten Wortgenerierungsaufgaben (*Determine best suited tasks*), Trainieren des Meta-Klassifikators (*Meta-classifier training*), Trainieren der Klassifikatoren (*Train classifiers*), Zusammenführen der Klassifikatoren (*Combine classifiers*); aus Thiele et al. (under revision, S. 24);

In der vorliegenden Studie wurden eine relativ simple (*Majority Voting*) und eine komplexere (*SVM*) Methode verwendet, um Klassifikatoren zu kombinieren und verschiedene Aufgaben zu evaluieren. Der relativ simple 1NN Klassifikator wurde verwendet, da vorausgehende Studien darauf hinwiesen, dass der 1NN Klassifikator vor dem Hintergrund des untersuchten Datensets eine gute Wahl sein würde (Gaspers et al., 2012). Ausgangspunkt für diese Herangehensweise war der Befund von Järvelin und Juhola (2011). Die Autoren verwendeten *machine learning*-Techniken, um Patienten mit Aphasie und Sprachgesunde zu klassifizieren und konnte zeigen, dass unterschiedliche Klassifikatoren für verschiedene Datensets die besten Leistungen

erbrachten. Die Autoren folgerten daraus, dass die Wahl eines geeigneten Klassifikators mit Bezug auf das verwendete Datenset getroffen werden sollte (Järvelin & Juhola, 2011). Ein Grund dafür, dass der simple 1NN Klassifikator in der vorliegenden Studie so gute Resultate lieferte, könnte darin liegen, dass das verwendete Datenset relativ klein war. Für andere und größere Datensets sind zukünftig komplexere Klassifikationsalgorithmen (wie z.B. *SVM*) zu bevorzugen. Mit dem Ziel langfristig die bestmögliche Strategie für die Klassifikation von Patienten mit kommunikativ-kognitiven Defiziten zu identifizieren, könnten auch noch weitere Meta-Klassifikatoren mit verschiedenen Architekturen, wie beispielsweise kaskadischen oder hierarchischen (vgl. Jain et al., 2000), angewendet und ihre Performanz miteinander verglichen werden.

Der Fokus der beschriebenen Studie bestand in der Analyse der Klassifikationsgenauigkeit als primäres Evaluationsmaß. Die in dieser Studie präsentierte Methodologie kann jedoch auch mit anderen Evaluationsmaßen angewendet werden, wie die Berücksichtigung der Maße *Precision* und *Recall* in Experiment 2b zeigte. Für die Entwicklung einer Methode, die darauf abzielt, eine reliable Bestätigung eines kommunikativ-kognitiven Defizits zu erreichen, könnte eine Evaluationsmethode, die auf das Maß der *Precision* innerhalb der Gruppe der Patienten mit kommunikativ-kognitiver Beeinträchtigung optimiert ist, einen geeigneten Ansatz liefern. So könnten die Berechnungen, die im Rahmen dieser Studie vorgestellt wurden, analog mit den Variablenkombinationen durchgeführt werden, für die die beste *Precision* in der Gruppe der Patienten mit kommunikativ-kognitiven Defiziten erzielt wurde. Dasselbe sollte anschließend auch für die Gruppe der Sprachgesunden durchgeführt werden, um das Nichtbestehen eines kommunikativ-kognitiven Defizits abzusichern. Die exakte Planung einer solchen Evaluationsmethode muss auf der Basis eines größeren Datensets stattfinden um in der klinischen Praxis anwendbar zu sein. Die beschriebenen Ergebnisse weisen jedoch auf die Möglichkeit der Entwicklung einer präzisen Methode hin, mit der ein Bestehen bzw. Nicht-Bestehen kommunikativ-kognitiver Defizite bestätigt werden könnte.

Eine weitere Erfolg versprechende Methode könnte sein, andere klinische Maße, welche z.B. kognitive Fähigkeiten wie etwa Arbeitsgedächtnisleistungen, *high-level-language*, Diskurs- oder Kommunikationsfähigkeiten erfassen, in die Analyse einzubeziehen. Defizite in diesen Domänen werden bei Patienten mit kommunikativ-kognitiven Beeinträchtigung beschrieben und führen zu den typischen, beobachtbaren klinischen Symptomen (Lé et al., 2011; Martin & McDonald, 2003; Prigatano, 1986). Die Leistungen in kognitiven, *high-level-language* und Diskursfähigkeiten könnten zusätzlich evaluiert werden, um anschließend die daraus resultierenden Ergebnisse mit denen aus den Wortgenerierungsleistungen zu kombinieren. Durch ein

Trainieren auf den Daten zusätzlicher Tests kognitiver sowie anderer Funktionen könnten Klassifikatoren gebildet werden die anschließend, parallel zu dem Vorgehen in dieser Studie, mit Klassifikatoren kombiniert werden, die auf der Basis von Wortgenerierungsdaten trainiert worden sind. Es wäre dann möglich, Klassifikatoren basierend auf den Ergebnissen aus mehreren solcher Tests gleichzeitig zu trainieren und anschließend die resultierenden Klassifikatoren mit denen aus den Wortgenerierungsaufgaben im Sinne einer Meta-Klassifikation zu kombinieren. Ein ähnliches Vorgehen beschreiben beispielsweise auch Quintana und Kollegen (2012), mit vielversprechenden Ergebnissen für die Klassifikation von Patienten mit leichten kognitiven Beeinträchtigungen und Patienten mit Demenz vom Alzheimer Typ.

Weiterhin könnte es aussichtsreich sein, die hier explorierten Klassifikationsmethoden anzuwenden, um andere Patientenpopulationen zu evaluieren und klassifizieren. Denkbar wäre etwa die Klassifikation von Patienten mit Multipler Sklerose, Autisten oder Patienten mit restaphasischer Symptomatik. Die beschriebene Methode könnte für all jene klinischen Populationen sinnvoll angewendet werden, für die eine reliable Diagnose mit standardisierten Tests noch eine Herausforderung darstellt bzw. für all diejenigen, für die noch keine geeignete Prozedur existiert. Das Einbeziehen anderer Aufgaben, Variablen und/oder Klassifikatoren in die beschriebene Evaluationsmethode könnte zur Fazilitierung der Diagnosestellung sinnvoll sein und gilt in weiteren Studien zu untersuchen. Für Sprachstörungen wie zum Beispiel die Aphasie, die zuverlässig mit standardisierten Verfahren und etablierten Methoden diagnostiziert werden kann (z.B. AAT, Huber et al., 1983) scheint bislang ein zusätzlicher Aufwand, der in der Analyse zusätzlicher Variablen besteht, nicht notwendig bzw. gerechtfertigt (Gaspers et al., 2012).

Finden Wortgenerierungsaufgaben in der klinischen Praxis Anwendung, werden Beurteilungen bezüglich *unterdurchschnittlicher* oder *durchschnittlicher* Leistungen auf der Basis der *Anzahl korrekter Nennungen*, des ermittelten Prozentrangs und dem Vergleich einer individuellen Leistung mit einer Referenzgruppe getroffen (Aschenbrenner et al., 2000). Die Anwendung und Evaluation von Wortgenerierungsaufgaben mit dieser traditionell angewendeten Prozedur ist schnell und unkompliziert. Jedoch, wie im Fall der beschriebenen Patienten mit kommunikativ-kognitiven Defiziten, kann die Klassifikation auf Basis von Prozenträngen schwierig und unbefriedigend sein, da sogar Sprachgesunde unterdurchschnittliche Leistungen und Patienten mit kommunikativ-kognitiven Defiziten häufig durchschnittliche oder sogar überdurchschnittliche Leistungen zeigen. Die Annotation und Evaluation zusätzlicher Variablen ist im Vergleich zur traditionellen klinischen Evaluation eher zeitaufwändig, da sie - basierend auf der Natur der Variablen die für die Evaluation von Wortgenerierungsleistungen sinnvoll scheinen - zu einem großen Teil die manuelle Annotation von zeitlichen, lexikali-

schen, sublexikalischen und/oder inhaltsbezogenen Variablen erfordert. Das Ergebnis einer besseren bzw. fazilitierten Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden scheint jedoch diesen Aufwand zu rechtfertigen.

Zum jetzigen Zeitpunkt ist die Evaluation zusätzlicher Variablen noch explorativ und muss aufgrund des geringen Datensets weiter validiert werden. Wenn die Variablen zur Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden auf der Basis eines größeren Datensatzes validiert sind, könnte es ausreichend sein, nur die für den Meta-Klassifikator relevanten Variablen zu extrahieren, um eine reliable Diagnose stellen zu können. Der Evaluationsaufwand für die klinische Praxis wäre dann wieder minimiert.

5.6.1 Fazit

Die Ergebnisse der drei Experimente weisen darauf hin, dass *machine learning*-Techniken die Identifizierung kommunikativ-kognitiver Defizite unterstützen können. Im Rahmen dieser Arbeit wurde die Möglichkeit der Entwicklung einer Evaluationsmethode auf Basis von Wortgenerierungsaufgaben exploriert. Bei der Evaluation von Wortgenerierungsleistungen wurden verschiedene Variablen und Variablenkombinationen identifiziert, die für die Diagnose kommunikativ-kognitiver Defizite und insbesondere für die Dissoziation zwischen Patienten mit kommunikativ-kognitiver Beeinträchtigung und Sprachgesunden gut geeignet sein könnten. Bei der Evaluation mit *machine learning*-Techniken zeigten diese Variablen und Variablenkombinationen bessere Klassifikationsergebnisse als das Maß *Anzahl korrekter Nennungen*, das traditionell in der klinischen Praxis bei der Evaluation von Wortgenerierungsleistungen angewendet wird.

Mit dem Einsatz von *machine learning*-Techniken bei der Diagnosestellung scheint sich eine Methodologie zu entwickeln, die, nach weiterer Validierung, die klinische Analyse zukünftig sinnvoll ergänzen kann.

Kapitel 6

Diskussion

Gegenstand der vorliegenden Dissertation war die Evaluation von Wortgenerierungsleistungen bei kommunikativ-kognitiven Defiziten. Es wurden drei Studien vorgestellt, die sich mit unterschiedlichen Methodologien bei der Analyse von Wortgenerierungsleistungen bei Patienten mit kommunikativ-kognitiven Defiziten, bei Patienten mit Aphasie sowie bei Patienten mit verschiedenen Verlaufsformen einer Multiplen Sklerose im Vergleich zu Sprachgesunden befassten. Das übergeordnete Ziel war dabei, einer reliablen Identifizierung und Diagnose kommunikativ-kognitiver Defizite unterschiedlicher Genese durch die Evaluation quantitativer sowie qualitativer Charakteristika in standardisierten (d.h. simple und komplexe lexikalische sowie semantische Wortgenerierungsaufgaben, Aschenbrenner et al., 2000) und (noch) nicht-standardisierten Aufgaben (d.h. Verbgenerierung, vgl. Piatt et al., 1999b oder Pekkala, 2012) näher zu kommen.

Spezifische Fragestellungen waren dabei:

- (a) Ist die Evaluation von Wortgenerierungsleistungen zur Diagnose kommunikativ-kognitiver Defizite und anderer Sprach- und Kommunikationsstörungen geeignet?
- (b) Welche Auswertungsparameter ermöglichen eine reliable Diagnosestellung? Unterscheiden sich die Auswertungsparameter für die Diagnose verschiedener Störungsbilder?
- (c) Welche Rückschlüsse können aus reduzierten Wortgenerierungsleistungen oder spezifischen Leistungsunterschieden gezogen werden?
- (d) Welche der verwendeten Methodologien kann für die Evaluation von Wortgenerierungsleistungen sinnvoll eingesetzt werden?

Der Beantwortung dieser Fragen (vgl. Abschnitt 6.2) geht im folgenden Abschnitt eine kurze Zusammenfassung der wichtigsten Ergebnisse aus den drei Studien voraus (siehe Abschnitt 6.1). Darauf aufbauend sollen im Anschluss in einem allgemeinen Fazit die Konsequenzen für das diagnostische Vorgehen aufgezeigt werden (siehe Abschnitt 6.3). Abschließend folgt ein Ausblick auf weitere Forschung zur Evaluation von Wortgenerierungsleistungen (siehe Abschnitt 6.4).

6.1 Zusammenfassung

Studie 1 *Evaluation quantitativer und qualitativer Charakteristika der Wortgenerierungsleistung bei Multipler Sklerose* (vgl. Kapitel 3) befasste sich mit der Evaluation quantitativer und qualitativer Charakteristika von Wortgenerierungsleistungen bei Patienten mit schubförmig-remittierender und chronisch-progredienter Multipler Sklerose im Vergleich zu Sprachgesunden. Neben einer quantitativen Analyse der *Anzahl korrekter Nennungen* wurden bei der qualitativen Auswertung einerseits der Anteil korrekter Nennungen in Relation zur Gesamtzahl der Nennungen (als indirektes Fehlermaß) sowie das Auftreten von Perseverationen und Regelbrüchen betrachtet. Das Ziel war dabei, der Charakterisierung eines möglichen sprachlichen und/oder kognitiv-exekutiven Leistungsdefizits bei Multipler Sklerose näher zu kommen.

Die Analysen ergaben schlechtere Wortgenerierungsleistungen für die Patienten mit Multipler Sklerose (unabhängig von der Verlaufsform) im Vergleich zu den Sprachgesunden. Insbesondere schlechtere Leistungen in den beiden Wechsel-Aufgaben wiesen auf ein unabhängig von der Verlaufsform der Erkrankung bestehendes Defizit in den Exekutivfunktionen hin. Eine Dissoziation zwischen den beiden Verlaufsformen der Multiplen Sklerose war anhand der Leistungen in den semantischen Wortgenerierungsaufgaben sowie bei der Verbgenerierung möglich. Dies wurde als Hinweis auf ein sich bei chronischer Progredienz der Erkrankung zunehmend entwickelndes Defizit im semantischen System gewertet. Für die Gruppe der Patienten mit chronisch-progredienter Multipler Sklerose wiesen Unterschiede bezüglich des Fehleranteils und des Anteils der Regelbrüche in der lexikalischen sowie bezüglich des Anteils der Kategorieperseverationen in der semantischen Wechsel-Aufgabe auf eine zusätzliche Progredienz exekutiver Defizite hin.

Die Ergebnisse dieser Studie legen nahe, dass bei Patienten mit Multipler Sklerose unabhängig von der Verlaufsform eine kognitiv-exekutiv bedingte Minderung der Wortgenerierungsleistung vorliegt. Bei progredientem Verlauf der Erkrankung verstärkt sich dieses exekutive Defizit und wird darüber hinaus von einem zunehmenden Defizit im semantischen System begleitet. Geht man davon aus, dass bei bis zu 80 Prozent der Erkrankten ein initial schubförmiger Verlauf in eine chronisch-progrediente Verlaufsform übergeht

(Poeck & Hacke, 2001; Pusswald & Vass, 2011), lässt sich eine Progredienz von relativ umschriebenen Defiziten hin zu globaleren Beeinträchtigungen erkennen. Dies wurde darüber hinaus durch die generell schlechteren kognitiven Leistungen der Patienten mit chronisch-progredienter Multipler Sklerose im DemTect (Kalbe et al., 2004) bestätigt.

In **Studie 2** *Evaluation von Wortgenerierungsleistungen zur Klassifikation von Sprach- und Kommunikationsstörungen mit machine learning-Techniken* (vgl. Kapitel 4) wurde untersucht, ob zusätzliche aus Wortgenerierungsdaten extrahierte Variablen dazu geeignet sind, zwischen Sprachgesunden und Patienten mit Aphasie sowie Patienten mit kommunikativ-kognitiven Defiziten zu dissoziieren. Es wurden verschiedene zeitliche, lexikalisch-semantische und sublexikalische Variablen mit *machine learning*-Techniken evaluiert, mit dem Ziel, das empirische Risiko einer Fehlklassifikation zu minimieren.

Die Analysen zeigten, dass das traditionell in der klinischen Praxis angewendete Maß *Anzahl korrekter Nennungen* gut geeignet ist, um zwischen Patienten mit Aphasie und Sprachgesunden sowie zwischen Patienten mit Aphasie und Patienten mit kommunikativ-kognitiven Defiziten zu unterscheiden. Bei der Dissoziation zwischen Patienten mit kommunikativ-kognitiven Beeinträchtigungen und Sprachgesunden konnten hingegen durch die Berücksichtigung zusätzlicher Variablen bessere Klassifikationsergebnisse im Vergleich zum traditionellen Maß erreicht werden: das beste Klassifikationsergebnis wurde erzielt, wenn die *Anzahl korrekter Nennungen* von der Analyse ausgeschlossen wurde. Darüber hinaus zeigten die Analysen, dass anhand der Leistungen in der simplen semantischen Wortgenerierungsaufgabe geringfügig bessere Klassifikationsergebnisse bei der Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden erzielt werden konnten. Eine anschließende Variablenanalyse ergab, dass insbesondere zeitliche Charakteristika (z.B. die kürzeste Zeitspanne zwischen den Startzeitpunkten konsekutiver Nennungen) aussichtsreiche Kandidaten für eine reliable Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden zu sein scheinen.

Die Ergebnisse dieser Studie legen nahe, dass die Evaluation zusätzlicher, qualitativer Auswertungsparameter eine sinnvolle Ergänzung traditioneller Analysemethoden darstellt, die die Identifizierung kommunikativ-kognitiver Defizite fazitätieren kann.

Studie 3 *Exploration einer Methodologie zur Evaluation von Wortgenerierungsleistungen mit machine learning-Techniken* (vgl. Kapitel 5) untersuchte die Möglichkeit der Entwicklung einer Evaluationsmethode zur Diagnose kommunikativ-kognitiver Beeinträchtigungen. Diese basierte auf der Anwendung von *machine learning*-Techniken auf Wortgenerierungsdaten von Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden.

Die Analysen der vorausgehenden Studie in Kapitel 4 wurden hierfür erweitert. Es wurden sowohl als vielversprechende Maße identifizierte Variablen der vorherigen Studie eingehender, als auch neue Variablen analysiert. Auf Basis dieser Variablen und einem Referenzdatenset wurden insgesamt drei Experimente durchgeführt. In Experiment 1 wurden Einzelvariablen sowohl mit *machine learning*-Techniken als auch mit klinisch-statistischen Analysen evaluiert. Mit Experiment 2 wurde der Frage nachgegangen, ob auf Variablenkombinationen trainierte Klassifikatoren besser für die Klassifikation geeignet sind als solche, die auf Einzelvariablen trainiert wurden. Neben der *accuracy* wurden hier weitere Maße (*Precision* und *Recall*) zur Beurteilung der Güte des Klassifikationsergebnisses eingesetzt. In Experiment 3 wurde anschließend exploriert, ob die Anwendung eines Meta-Klassifikators, der die Ergebnisse individueller auf aufgabenspezifischen Variablenkombinationen trainierter Klassifikatoren miteinander kombiniert, die Klassifikationsgenauigkeit weiter verbessert.

Übereinstimmend mit den Ergebnissen aus Studie 2 (vgl. Kapitel 4) wiesen die Ergebnisse dieser Studie darauf hin, dass (i) die *Anzahl korrekter Nennungen*, das Maß das traditionell in der klinischen Praxis zur Evaluation von Wortgenerierungsleistungen verwendet wird, nicht hinreichend gut geeignet zu sein scheint, um kommunikativ-kognitive Defizite reliabel zu diagnostizieren. Die Ergebnisse zeigten auch, dass (ii) durch die Anwendung aufgabenspezifischer Variablen oder Variablenkombinationen bei der Evaluation der einzelnen Wortgenerierungsaufgaben bessere Ergebnisse erzielt werden können als bei Anwendung derselben Variablen oder Variablenkombinationen gemittelt über alle Aufgaben. Des Weiteren zeigte sich, dass (iii) Variablenkombinationen besser für die Dissoziation geeignet zu sein scheinen als Einzelvariablen und dass (iv) das Trainieren von Klassifikatoren auf verschiedenen Wortgenerierungsaufgaben mit aufgabenspezifischen Variablenkombinationen und eine anschließende Meta-Klassifikation die Performanz der Klassifikation weiter verbessert.

Während mit der *Anzahl korrekter Nennungen* gemittelt über alle Aufgaben eine Klassifikationsgenauigkeit erzielt wurde, die noch unterhalb der einer Zufallsklassifikation lag, konnten bei Anwendung des Meta-Klassifikators nahezu alle Patienten korrekt klassifiziert werden. Wie auch in der vorausgehenden Studie, zeigte sich auch in dieser eine zeitliche Variable (d.h. die kürzeste Pause zwischen konsekutiven Nennungen) als geeigneter Kandidat für die Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden. Darüber hinaus waren insbesondere bei der Evaluation von Einzelvariablen Hinweise darauf zu finden, dass Leistungen in semantischen Wortgenerierungsaufgaben besser zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden unterscheiden können.

Insgesamt zeigen die Ergebnisse dieser Studie, dass die Verwendung von Wortgenerierungsdaten in Kombination mit *machine learning*-Techniken ein

vielversprechender Ansatz für die Entwicklung einer reliablen Evaluationsmethode zur Diagnose kommunikativ-kognitiver Defizite zu sein scheint.

Zusammengefasst sind die wichtigsten Befunde aus den drei Studien: (i) Die *Anzahl korrekter Nennungen* scheint nicht für alle der untersuchten Störungsbilder ein geeignetes Maß zur Identifizierung eines Leistungsdefizits darzustellen; (ii) die zusätzliche Berücksichtigung qualitativer Aspekte scheint eine sinnvolle Ergänzung bei der Evaluation von Wortgenerierungsleistungen zu sein, insbesondere bei der Identifizierung vergleichsweise subtiler Defizite, wie etwa bei der Diagnose kommunikativ-kognitiver Beeinträchtigungen; (iii) Leistungen in semantischen Wortgenerierungsaufgaben scheinen besser zwischen spezifischen Populationen zu dissoziieren als Leistungen in lexikalischen Aufgaben; (iv) mit *machine learning*-Techniken ist eine reliable Dissoziation zwischen Patienten mit Sprach- und/oder Kommunikationsstörungen unterschiedlicher Genese und Sprachgesunden auf Basis von Wortgenerierungsdaten möglich.

Die beschriebenen Befunde werden nun vor dem Hintergrund diagnostischer Implikationen diskutiert.

6.2 Evaluation von Wortgenerierungsleistungen

Lassen sich verschiedene klinische oder nicht-klinische Populationen voneinander abgrenzen, ob anhand der *Anzahl korrekter Nennungen*, verschiedener qualitativer Charakteristika oder Leistungen in einer spezifischen Wortgenerierungsaufgabe, kann das als Hinweis auf spezifische defizitäre Leistungen in zugrunde liegenden Kompetenzen gesehen werden. Daraus ergibt sich die Frage nach der Bedeutung dieser Unterschiede und den daraus für das diagnostische und therapeutische Vorgehen resultierenden Konsequenzen.

In den nun folgenden Abschnitten sollen die beschriebenen Ergebnisse studienübergreifend diskutiert werden. Im Vordergrund stehen dabei zunächst Implikationen bezüglich der Relevanz quantitativer und qualitativer Auswertungsparameter, wobei insbesondere die Rolle zeitlicher Charakteristika bei der Evaluation von Wortgenerierungsaufgaben fokussiert werden soll. Im Weiteren wird dann die Evaluation von Leistungen in verschiedenen Wortgenerierungsaufgaben näher betrachtet. Anschließend folgt die Darstellung eines Einzelfalls, der sozial-kognitive Aspekte bei der Evaluation von Wortgenerierungsleistungen beleuchtet. Darüber hinaus sollen Aspekte der Anwendbarkeit von *machine learning*-Techniken bei der Evaluation von Wortgenerierungsleistungen diskutiert werden.

6.2.1 Evaluation quantitativer und qualitativer Auswertungsparameter: Zur Bedeutung zeitlicher Charakteristika

Im Mittelpunkt der beschriebenen Studien stand die Evaluation von Wortgenerierungsleistungen in verschiedenen Populationen unter Berücksichtigung sowohl quantitativer als auch qualitativer Auswertungsparameter.

Traditionell wird bei der Evaluation von Wortgenerierungsleistungen die *Anzahl korrekter Nennungen* als Maß zur Leistungsbeurteilung verwendet (Aschenbrenner et al., 2000; Paulsen et al., 1996). Die *Anzahl korrekter Nennungen* spiegelt als klassisch quantitatives Maß die Korrektheit der Ausführung einer Aufgabe wider; eine reduzierte *Anzahl korrekter Nennungen* entspricht einer schlechteren Leistung bei der Wortgenerierung. Wodurch diese reduzierte Wortgenerierungsleistung bedingt ist, geht aus der *Anzahl korrekter Nennungen* nicht hervor (Drechsler, 2007).

Eines der interessantesten und wichtigsten Ergebnisse dieser Arbeit ist, dass die Evaluation dieses traditionellen Maßes nicht bei allen untersuchten klinischen Populationen zur Identifizierung eines Leistungsdefizits geeignet war. Sowohl in Studie 2 (vgl. Kapitel 4) bei der Dissoziation zwischen Sprachgesunden, Aphasikern und Patienten mit kommunikativ-kognitiven Defiziten, als auch in Studie 3 (vgl. Kapitel 5) bei der die Exploration einer Methodologie zur reliablen Identifizierung von kommunikativ-kognitiven Defiziten im Vordergrund stand, konnten Hinweise darauf gefunden werden, dass Patienten mit kommunikativ-kognitiven Defiziten besser mit anderen als der traditionell zur Evaluation verwendeten Variable von Sprachgesunden zu unterscheiden sind. Für Studie 3 ist darüber hinaus zu betonen, dass auch mit statistischen Analysen in keiner der vier Wortgenerierungsaufgaben signifikante Unterschiede bezüglich der *Anzahl korrekter Nennungen* zwischen Sprachgesunden und Patienten mit kommunikativ-kognitiven Defiziten gefunden werden konnten. Darüber hinaus war zu beobachten, dass die Klassifikationsgenauigkeit des traditionellen Maßes bei der Evaluation mit *machine learning*-Techniken (vgl. Experiment 1b in Studie 3, Abschnitt 5.3.2) stets unterhalb der lag, die bei einer zufälligen Klassifikation erreicht werden könnte. Dies kann als ein Indiz dafür gesehen werden, dass die Analyse der *Anzahl korrekter Nennungen* die Identifizierung von kommunikativ-kognitiven Defiziten nicht faziilitiert, sondern vielmehr zu erschweren scheint.

Die Feststellung, dass die Untersuchung quantitativer Aspekte einer Leistung nicht immer zur Abgrenzung verschiedener Störungsbilder und insbesondere zur Identifizierung relativ subtiler Beeinträchtigungen beitragen, beschreiben auch andere Autoren (vgl. hierzu beispielsweise Possin et al., 2005, für die Dissoziation zwischen kortikaler und subkortikaler Demenz oder Troyer & Moscovitch, 2006). Was bleibt ist die Frage, warum ein quantitatives Maß wie die *Anzahl korrekter Nennungen* nicht zur Abgrenzung

sprachgesunder Leistungen von kommunikativ-kognitiven Defiziten geeignet ist, während es sehr wohl zur Dissoziation anderer Störungsbilder (z.B. Aphasie vs. Sprachgesund) reliabel beitragen kann?

Während die *Anzahl korrekter Nennungen* als kein geeignetes Maß schien, um Patienten mit kommunikativ-kognitiven Defiziten zu identifizieren, konnte es unter anderem reliabel zwischen Patienten mit Aphasie und Sprachgesunden sowie zwischen Patienten mit Aphasie und Patienten mit kommunikativ-kognitiven Defiziten differenzieren (vgl. Studie 2, Kapitel 4). Dieses Ergebnis war nicht weiter verwunderlich, da aphasische Sprachproduktion durch lexikalische und semantische Verarbeitungsdefizite in Speicherung, Enkodierung und Abruf gekennzeichnet ist (Adams et al., 1989; Basso et al., 1997; Davis, 2007; Grossman, 1981). Beeinträchtigungen in der Wortgenerierung werden demnach häufig beschrieben und sollen vor allem durch eine geringere *Anzahl korrekter Nennungen* hervortreten (Basso et al., 1997). Es war daher zu erwarten, dass sich Aphasien anhand der *Anzahl korrekter Nennungen* adäquat klassifizieren lassen.

Darüber hinaus unterschied sich die *Anzahl korrekter Nennungen* bei den Sprachgesunden im Vergleich zu Patienten mit chronisch-progredienter und Patienten mit schubförmig-remittierender Multipler Sklerose. Dabei war ein treppenartiges Leistungsprofil zu erkennen: Die *Anzahl korrekter Nennungen* zeigte sich bei den Sprachgesunden am höchsten, gefolgt von den Patienten mit schubförmig-remittierender Multipler Sklerose, die wiederum mehr korrekte Nennungen produzierten als Patienten mit chronisch-progredienter Multipler Sklerose. Dieses Ergebnis war zu erwarten, insbesondere deshalb, da Wortgenerierungsleistungen als sehr sensibles Maß zur Identifizierung von Leistungsdefiziten bei Multipler Sklerose beschrieben werden (vgl. Henry & Beatty, 2006). Darüber hinaus wird häufig, wenn auch kontrovers, von schlechteren Leistungen bei Patienten mit chronisch-progredienter Multipler Sklerose im Vergleich zu Patienten mit schubförmig-remittierendem Verlauf der Erkrankung berichtet (Henry & Beatty, 2006; Zakzanis, 2000; Messinis et al., 2013). Interessant wird das Ergebnis, wenn man sich vor Augen führt, welche sprachlichen bzw. kommunikativ-kognitiven Defizite im Rahmen einer Multiplen Sklerose häufig beschrieben werden. Hierzu zählen insbesondere Defizite in der *high level language*, welche sich unter anderem in Beeinträchtigungen beim Schlussfolgern oder Verstehen mehrdeutiger Äußerungen zeigen können oder sich in der Spontansprache durch einen Mangel an Kohärenz ausdrücken (vgl. Arnott et al., 1997; Laakso et al., 2000; Lethlean & Murdoch, 1997). Wie in Abschnitt 2.4.3 beschrieben, ähneln die im Rahmen einer Multiplen Sklerose beobachtbaren Defizite häufig denen kommunikativ-kognitiver Beeinträchtigungen und weniger den Symptomen einer sprachsystematischen Einschränkung im Sinne einer Aphasie (Lacour et al., 2004). Auch läsionslokalisationstheoretisch lassen sich diesbezüglich Zusammenhänge zwischen kommunikativ-kognitiven Defiziten und Sprach-

und Kommunikationsstörungen im Rahmen einer Multiplen Sklerose erkennen. So werden kommunikativ-kognitive Defizite vor allem nach rechts-hemisphärischen oder (prä)frontalen, aber auch nach subkortikalen Läsionen berichtet (Blake et al., 2002; Cummings, 2007; Martin & McDonald, 2003). Demgegenüber ist die Multiple Sklerose zwar durch eine große Heterogenität möglicher Läsionslokalisationen geprägt (Pusswald & Vass, 2011), vor allem bei Progredienz der Erkrankung wird jedoch von zunehmenden Demyelinisierungsherden in der grauen Substanz sowie insbesondere in frontalen Arealen berichtet (Geurts et al., 2012; Kraus et al., 2005; Zakzanis, 2000). Es lassen sich demnach Überschneidungen bei den Läsionslokalisationen im Rahmen einer Multiplen Sklerose und bei kommunikativ-kognitiven Defiziten anderer Ätiologien erkennen.

Die Feststellung, dass sich Patienten mit kommunikativ-kognitiven Defiziten und Patienten mit Multipler Sklerose hinsichtlich ihrer sprachlich-kommunikativen Leistungen ähneln, jedoch Unterschiede in der Wortgenerierungsfähigkeit aufzeigen, kann ein Indiz dafür sein, dass der beobachtbaren sprachlichen bzw. kommunikativ-kognitiven Performanz unterschiedliche, defizitäre Kompetenzen zugrunde liegen (Lezak et al., 2004). So könnten beispielsweise Beeinträchtigungen in unterschiedlichen Komponenten von Exekutivfunktionen, das heißt defizitäre Funktionen der *Initiierung*, der *Inhibition*, des *Monitoring*, der *Strategieanwendung* (inkl. *set-shifting*) oder der zeitlichen Integration kognitiver und sprachlicher Verarbeitungsprozesse, zu ähnlichen Auffälligkeiten in Spontansprache und Kommunikation führen.

Wenn traditionelle Maße zur Identifizierung kommunikativ-kognitiver Defizite nicht hinreichend geeignet sind, stellt sich die Frage welche Variablen oder Auswertungsparameter anstelle der *Anzahl korrekter Nennungen* eine reliable Diagnosestellung ermöglichen können?

Es zeigte sich, dass insbesondere Variablen, die auf der Analyse zeitlicher Charakteristika beruhen (wie z.B. die kürzeste Pause zwischen konsekutiven Nennungen), als geeignete Kandidaten für die Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden identifiziert werden konnten. So wies das Wortgenerierungsprofil sprachgesunder Versuchsteilnehmer im Vergleich zu Patienten mit kommunikativ-kognitiven Defiziten beispielsweise eine signifikant kürzere kürzeste Pause zwischen konsekutiven Nennungen auf. Aber auch andere zeitliche Charakteristika wie etwa die durchschnittliche Pausenlänge traten hervor. Studie 2 konnte darüber hinaus mit der Variablenanalyse (vgl. Abschnitt 4.3.2) zeigen, dass zeitbasierte Variablen bei allen Klassifikationsaufgaben und nicht nur bei der Dissoziation zwischen sprachgesunden Leistungen und kommunikativ-kognitiven Defiziten, stark vertreten waren. Das Besondere war, dass zeitbasierte Auswertungsparameter bei Patienten mit kommunikativ-kognitiven Defiziten besser zur Identifizierung des Defizits beitragen konnten. Die Frage, die sich aus den beschriebenen Befunden ergibt, ist: Warum sind gerade

zeitliche Charakteristika sensitiv für die Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden?

Verschiedene Theorien exekutiver Funktionen (z.B. Baddeley 2003a; Calahan et al., 2001; Fuster, 2002, Lezak et al., 2004 oder Smith & Jonides, 1999; siehe auch Abschnitt 2.2.3) stellen den Prozess der zeitlichen Integration als wichtige Komponente von Exekutivfunktionen heraus. Nach Fuster (2002) besteht die zeitliche Integration kognitiv-exekutiver Prozesse in der Umsetzung eines spezifischen Inputs in einen kohärenten Output. Die Koordination des Zusammenspiels verschiedener relevanter kognitiv-exekutiver Funktionen ist dabei von besonderer Bedeutung. Smith und Jonides (1999) beschreiben die zeitliche Integration als Prozesskomponente, der vor allem bei komplexen Aufgaben, die einen Wechsel des Aufmerksamkeitsfokus verlangen, eine zentrale Rolle bei der Steuerung der Aufgabenbewältigung zukommt. Und auch Hermann und Grabowski (1994) betonen in ihrer sprachpsychologisch ausgerichteten Theorie der Sprachverarbeitung (Mannheimer Regulationstheorie; vgl. Kapitel 2, Abschnitt 2.1.1) die Rolle der *Zentralen Exekutive* bei der prozeduralen Organisation. Die Aufgabe der *Zentralen Exekutive* besteht dabei im Wesentlichen in der Selektion, Aufbereitung und Linearisierung von Informationen, die sie aus einem Arbeitsspeicher erhält. Insbesondere das Linearisieren von Information erfordert eine adäquate zeitliche Integration (Herrmann, 2003).

Dass vor allem zeitliche Variablen gut geeignet zu sein scheinen, um zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden zu dissoziieren, könnte darauf hinweisen, dass die zeitliche Integration verschiedener sprachlicher und nicht-sprachlicher kognitiver Funktionen, die für eine erfolgreiche Lösung von Wortgenerierungsaufgaben erforderlich sind, bei diesen Patienten beeinträchtigt ist. So könnte man annehmen, dass basale Komponenten kognitiv-exekutiver Funktionen, wie *Initiierung*, *Inhibition*, *Monitoring* und *Strategieranwendung* (inkl. *set-shifting*), bei Patienten mit kommunikativ-kognitiven Defiziten intakt sind, während die zeitliche Integration ebendieser Komponenten sowie lexikalischer und semantischer Verarbeitungsschritte defizitär ist.

Auch die Ergebnisse anderer Studien stützen diese Annahme einer defizitären zeitlichen Integration und bestätigen die Relevanz zeitlicher Charakteristika bei der Identifizierung kommunikativ-kognitiver Defizite (Barrow et al., 2003; Barrow et al., 2006; King, Hough, Walker, Rastatter, & Holbert, 2006a; King, Hough, Vos, Walker, & Givens, 2006b). So fanden Barrow und Kollegen (2003; 2006) bei der Untersuchung von Patienten mit minimalem Schädelhirntrauma Hinweise darauf, dass zeitliche Maße, insbesondere zur Identifizierung relativ geringer und standarddiagnostisch nicht erfassbarer Beeinträchtigungen, besser geeignet sind als traditionelle Evaluationsmaße, die in erster Linie die Korrektheit einer Leistung abbilden. Die Autoren un-

tersuchten Leistungen beim Benennen unter Bedingungen mit erhöhten kognitiven Anforderungen. Anstelle einer klassischen konfrontativen Benennungsaufgabe wurde mit den Teilnehmern eine "within category naming" Aufgabe durchgeführt. Das heißt, anstelle des abgebildeten Objektes sollte innerhalb kürzester Zeit ein anderer Vertreter aus der Kategorie des Objektes genannt werden (Beispiel: *Objekt = Apfel* → *Nennung = Banane*; Barrow et al., 2003; Barrow et al., 2006). Zur Identifizierung von subtilen Beeinträchtigungen wurden neben Maßen, die die Korrektheit einer Leistung abbilden, zusätzlich Reaktionszeitmessung eingesetzt. So konnten Barrow und Kollegen (2003; 2006) zeigen, dass zeitliche Charakteristika besser geeignet sind, um Patienten mit kommunikativ-kognitiven Defiziten nach minimalem Schädelhirntrauma zu identifizieren. Dies wurde darüber hinaus durch eine fehlende Korrelationen der erhobenen Reaktionszeiten und Korrektheitsmaße mit dem SCATBI (*Scales of Traumatic Brain Injury*, Adamovich & Henderson, 1992), einem etablierten Verfahren zur Erfassung kognitiver Leistungen nach Schädelhirntrauma, betont. Die Autoren sehen dies als weiteres Indiz dafür, dass zeitliche Maße auch solche Beeinträchtigungen aufdecken können, die mit traditionellen Verfahren nicht erfasst werden können (Barrow et al., 2006). Die Ursache für die Sensitivität zeitlicher Charakteristika sahen die Autoren in einer eingeschränkten kognitiven Flexibilität, das heißt einer defizitären zeitlichen Integration, welche sich auf die Geschwindigkeit der Bearbeitung einer Aufgabe auswirken kann. Eine reduzierte Verarbeitungsgeschwindigkeit wird bei minimalem Schädelhirntrauma häufig beschrieben und unter anderem neuropathologisch mit axonalen Schäden in Verbindung gebracht (Barrow et al., 2006).

Interessant wäre nun, zeitliche Charakteristika auch bei der Evaluation von Wortgenerierungsleistungen bei Patienten mit Multipler Sklerose zu berücksichtigen. Unterschiede in zeitlichen Parametern sind bei dieser klinischen Population zu erwarten, da eine reduzierte Informationsverarbeitungsgeschwindigkeit zu den am häufigsten beschriebenen kognitiven Beeinträchtigungen bei Patienten mit Multipler Sklerose gehört (Chiaravalloti & Luca, 2008; Pusswald & Vass, 2011; Rao et al., 1989). Auch neuroanatomisch bzw. -pathologisch wären Unterschiede zeitlicher Variablen plausibel, da die im Rahmen einer Multiplen Sklerose typische Demyelinisierung zu einer verlangsamten Reizweiterleitung führt (Pinel & Pauli, 2007), welche sich nachfolgend unter anderem in einer verlangsamten Informationsverarbeitungsgeschwindigkeit und/oder einer eingeschränkten kognitiven Flexibilität äußert (Pusswald & Vass, 2011).

An dieser Stelle lässt sich vermuten, dass bei vergleichsweise geringen Beeinträchtigungen, wie häufig im Falle kommunikativ-kognitiver Defizite, die relevanten Verarbeitungskomponenten (d.h. lexikalisch-semantisch, exekutive Basiskomponenten) intakt sind, aber das koordinierte Zusammenspiel nicht adäquat gelingt, was sich in abweichenden zeitlichen Charakte-

ristika bei der Wortgenerierung zeigt. Eine reduzierte Wortgenerierungsleistung in Form einer geringeren *Anzahl korrekter Nennungen* wäre demnach eher bei Defiziten in den beteiligten Verarbeitungskomponenten selbst zu erwarten (d.h. kognitiv-exekutiv oder lexikalisch/semantisch). Während bei kommunikativ-kognitiven Beeinträchtigungen Defizite bei der zeitlichen Integration relevanter Verarbeitungskomponenten zu Unterschieden zeitlicher Charakteristika führen, können demgegenüber im Rahmen einer Aphasie Defizite in der lexikalisch-semantischen Verarbeitung als Ursache für die reduzierte Wortgenerierungsleistung eine naheliegende Erklärung liefern (Baldo et al., 2010). Worauf die reduzierten Wortgenerierungsleistungen bei Multipler Sklerose zurückzuführen sein könnten, ist weniger einfach zu beantworten. Sowohl Defizite bei der lexikalisch-semantischen Verarbeitung oder Beeinträchtigungen in spezifischen kognitiv-exekutiven Komponenten sind hier denkbar. Weitere Untersuchungen, auch unter Einbezug zeitlicher Charakteristika bei der Wortgenerierung, sind hier erforderlich um die formulierten Annahmen weiter zu untersuchen.

Vor dem Hintergrund der Fragestellungen dieser Dissertation lässt sich an dieser Stelle ein Fazit ziehen. So scheint die Evaluation von Wortgenerierungsleistungen zur Identifizierung verschiedener Störungsbilder geeignet. Um auch vergleichsweise subtile Defizite identifizieren zu können, sollten zusätzlich zum traditionell verwendeten Evaluationsmaß *Anzahl korrekter Nennungen* weitere qualitative Charakteristika analysiert werden. Für die Diagnose kommunikativ-kognitiver Defizite scheinen vor allem zeitliche Parameter (z.B. die kürzeste oder durchschnittliche Pause zwischen zwei Nennungen) als geeignet.

Der folgende Abschnitt befasst sich nun mit der Frage nach der Bedeutung von Leistungsunterschieden in den verschiedenen Wortgenerierungsaufgaben.

6.2.2 Evaluation von Leistungen in unterschiedlichen Wortgenerierungsaufgaben

Ein weiteres interessantes und zugleich wichtiges Ergebnis stellt der Befund dar, dass Leistungen in semantischen Wortgenerierungsaufgaben besser zwischen spezifischen Patientengruppen zu differenzieren scheinen als Leistungen in lexikalischen Wortgenerierungsaufgaben. Dies trifft insbesondere für die Dissoziation zwischen Patienten mit schubförmig-remittierender und Patienten mit chronisch-progredienter Multipler Sklerose in Studie 1 (vgl. Kapitel 3), aber auch für die Klassifikation von Patienten mit kommunikativ-kognitiven Defiziten im Vergleich zu Sprachgesunden in Studie 2 (vgl. Kapitel 4) und Studie 3 (vgl. Kapitel 5) zu. In Studie 1 konnten anhand der Leistungen in semantischen Wortgenerierungsaufgaben nicht nur Sprachgesun-

de und Patienten mit Multipler Sklerose unabhängig von der Verlaufsform voneinander unterschieden, sondern auch zwischen den beiden Verlaufsformen der Multiplen Sklerose dissoziiert werden. In Studie 2 zeigten sich bei der Klassifikation von Sprachgesunden und Patienten mit kommunikativ-kognitiven Defiziten die besten Ergebnisse für die simple semantische Wortgenerierungsaufgabe (siehe Tabelle 4.9). In Studie 3 waren bei der Evaluation von Einzelvariablen in den verschiedenen Wortgenerierungsaufgaben (Experiment 1b; Abschnitt 5.3.1) ebenfalls bessere Klassifikationsergebnisse für die beiden semantischen Aufgaben zu erkennen. Dabei war die Klassifikationsgenauigkeit mit der simplen semantischen Aufgabe im Vergleich zur semantischen Wechsel-Aufgabe geringfügig höher.

Nach Henry und Beatty (2006) weisen größere Defizite in semantischen im Vergleich zu lexikalischen Wortgenerierungsaufgaben auf ein semantisches Defizit hin. Übertragen heißt das, wenn semantische Aufgaben besser zwischen spezifischen Populationen differenzieren können, kann eine Beeinträchtigung der semantischen Verarbeitung (zumindest anteilig) als Ursache angenommen werden. Die bessere Dissoziation anhand von Leistungen in semantischen Wortgenerierungsaufgaben zeigt an, dass zwischen den relevanten Populationen Unterschiede in den für die semantische Wortgenerierung erforderlichen Kompetenzen existieren; diese Unterschiede machen die Abgrenzung möglich. Das bedeutet auch, dass die Kompetenzen als defizitär angenommen werden können, die für eine erfolgreiche Lösung semantischer im Gegensatz zu lexikalischer Aufgaben erforderlich sind.

Während lexikalische Wortgenerierungsleistungen vor allem mit Funktionen im links-frontalen Kortex in Zusammenhang gebracht werden, sind semantische Wortgenerierungsleistungen mit der Integrität des links-temporalen Kortex assoziiert (Baldo et al., 2001, 2010; Pekkala, 2012). Bildgebungsstudien konnten darüber hinaus zeigen, dass semantische Verarbeitung auf einem weit verzweigten Netzwerk spezialisierter Funktionen basiert, die neuroanatomisch vergleichsweise weit repräsentiert sind und große Teile des linken Kortex einnehmen (Binder, Desai, Graves, & Conant, 2009). Insbesondere die semantische Enkodierung wird im Zusammenhang mit dem präfrontalen Kortex beschrieben (Perani et al., 2003). Dem linken inferioren frontalen Gyrus (IFG) soll dabei eine besondere Rolle zukommen. Während einige Autoren die Rolle des linken IFG in der semantischen Verarbeitung und dem Abruf semantischen Wissens sehen, argumentieren Perani und Kollegen (2003) "that it is not the retrieval of semantic knowledge per se that is associated with left IFG activation, but rather the need to select some relevant feature of semantic knowledge from a set of competing alternatives" (Perani et al., 2003, S. 365). Die Autoren nehmen an, dass anstelle des Abrufs semantischen Wissens, vielmehr die strategische Suche bzw. Auswahl relevanter semantischer Charakteristika funktionell im frontalen Kortex lo-

kalisiert ist (Perani et al., 2003).

Sind Funktionen neuroanatomisch vergleichsweise weit repräsentiert, wie es für die semantische Verarbeitung angenommen werden kann (Baldo et al., 2001, 2010; Pekkala, 2012; Perani et al., 2003), können selbst diffuse Läsionen zu Beeinträchtigungen führen. Auch King und Kollegen (2006a) beschreiben, dass sowohl fokale als auch diffuse Hirnschädigungen zu Beeinträchtigungen der Verarbeitung im semantischen System führen (King et al., 2006a). Die Evaluation von Leistungen in semantischen Wortgenerierungsaufgaben bietet demnach gegenüber der Analyse lexikalischer Wortgenerierungsleistungen die Möglichkeit, Defizite zu identifizieren, die aus einem vergleichsweise großen Areal möglicher Läsionslokalisationen resultieren. Insbesondere semantische Wortgenerierungsaufgaben sollten daher bei Störungen verschiedener Ätiologien und den für sie typischen Läsionslokalisationen sinnvoll einsetzbar sein, um Leistungsdefizite zu identifizieren. Das zeigt sich auch in den drei Studien, die im Rahmen dieser Dissertation vorgestellt wurden. Anhand der Leistungen in semantischen Wortgenerierungsaufgaben konnte nicht nur zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden sowie zwischen Patienten mit Multipler Sklerose unterschiedlicher Verlaufsformen reliabel unterschieden werden. Auch bei der Dissoziation zwischen Patienten mit Aphasie und Sprachgesunden in Studie 2 (vgl. Kapitel 4) wurden mit den semantischen Aufgaben sehr gute Klassifikationsergebnisse erzielt.

In Anlehnung an die Aussage von Perani und Kollegen (2003) stellt sich jedoch auch die Frage, ob die bessere Dissoziation anhand semantischer Wortgenerierungsleistungen auf beeinträchtigte Abrufprozesse oder strategisch-exekutive Prozesse bei der semantischen Wortgenerierung zurückzuführen sind. Dies ist mit den vorliegenden Daten nicht einfach zu beantworten. Es lassen sich jedoch vorsichtige Vermutungen formulieren.

Für die in Studie 2 und 3 untersuchten Patienten mit kommunikativ-kognitivem Defizit scheinen Beeinträchtigungen kognitiv-exekutiver Prozesse bei der semantischen Verarbeitung wahrscheinlicher. In Übereinstimmung mit Befunden aus anderen Studien (Barrow et al., 2003, 2006; King et al., 2006b,a) wurde gezeigt, dass ein Korrektheitsmaß weniger gut zur Identifizierung kommunikativ-kognitiver Defizite geeignet scheint als die Evaluation zeitlicher Charakteristika. Insbesondere in semantischen Wortgenerierungsaufgaben konnten mit den zeitbasierten Variablen gute Klassifikationsergebnisse erzielt werden (siehe Experiment 1b in Kapitel 5, vgl. Tabelle 5.3), was ein Hinweis darauf sein könnte, dass die kognitive Flexibilität vor allem im Hinblick auf die Verarbeitung und Integration semantischer Information beeinträchtigt ist. Auch die von Kavé und Kollegen (2011) beschriebenen Ergebnisse weisen in eine ähnliche Richtung. Die Autoren fanden bei einer Analyse lexikalischer und semantischer Wortgenerierungsleistungen bei

Patienten mit Schädelhirntrauma ebenfalls schlechtere Leistungen in den semantischen Aufgaben und interpretierten diese als kognitiv-exekutiv bedingtes Defizit. Gegenstand der Studie von Kavé et al. (2011) war die Frage, welche Art von Wortgenerierungsaufgabe (lexikalisch oder semantisch) für die Identifizierung kognitiv-exekutiver Defizite bei Patienten mit Schädelhirntrauma sensitiver ist. Die Analyse zeigte, dass sich insbesondere die Anzahl der *Switches* zwischen semantischen Subkategorien für die Gruppe der Schädelhirntraumapatienten unterschied. *Switches* zwischen Subkategorien werden, wenn auch kontrovers, als strategische Funktion beschrieben, die für eine erfolgreiche Lösung von Wortgenerierungsaufgaben erforderlich ist (Troyer et al., 1997; Troyer & Moscovitch, 2006). Dementsprechend interpretieren auch Kavé und Kollegen (2011) die bessere Dissoziation anhand von semantischen Wortgenerierungsaufgaben. Die Autoren gehen davon aus, dass sich semantische Aufgaben deshalb als besonders sensitiv zeigten, weil die relevante, defizitäre Komponente (Anzahl *Switches*) Ausdruck kognitiv-exekutiver Prozesse ist.

Für die in Studie 1 (vgl. Kapitel 3) untersuchten Patienten mit Multipler Sklerose ist eine spezifische Abgrenzung nicht möglich. Beeinträchtigungen bei der lexikalischen und semantischen Wortgenerierung lagen für beide Verlaufsformen im Vergleich zur Gruppe sprachgesunder Versuchsteilnehmer vor. Daher ist sowohl von Defiziten in den Exekutivfunktionen als auch von Beeinträchtigungen in der Semantik auszugehen. Welchen Anteil eine Beeinträchtigung semantischer Abrufprozesse oder strategisch-exekutive Defizite dabei haben, lässt sich auf Basis der vorliegenden Analysen und Ergebnisse nicht klären. Auch anhand von Befunden anderer Studien lässt sich keine sichere Aussage treffen. Multiple Sklerose ist generell durch ein sehr heterogenes Bild möglicher Symptome charakterisiert. Neben Defiziten bei der lexikalischen Verarbeitung (Sepulcre et al., 2011) wird von semantischen Zugriffstörungen wie auch von defizitär organisierten semantischen Netzwerken berichtet (Drake et al., 2006; Lethlean & Murdoch, 1994). Auch Defizite in den Exekutivfunktionen werden häufig beschrieben (Butters et al., 1998; Chiaravalloti & Luca, 2008; Pusswald & Vass, 2011). Weitere Untersuchungen sind hier erforderlich um weitergehende Aussagen machen zu können.

Ein weiterer Aspekt ist bei der Betrachtung semantischer Wortgenerierungsleistungen relevant. Semantische Wortgenerierungsaufgaben werden auf der Basis von semantischen Suchstrategien gelöst (Basso et al., 1997). Diese sollen (im Gegensatz zur lexikalischen/phonologischen Suche bei lexikalischen Wortgenerierungsaufgaben) der Strategie entsprechen, die auch in der alltäglichen Kommunikation bei der "Wortsuche" angewendet wird (Basso et al., 1997). Geht man davon aus, dass semantische Aufgaben eine etablierte semantische Suchstrategie erfordern, lässt sich vermuten, dass sich semantische Defizite bzw. Unterschiede bezüglich semantischer Parameter im Vergleich zu anderen relevanten Populationen auch in der Spontanspra-

che zeigen sollten. King und Kollegen (2006a) untersuchten den Wortabruf bei Patienten mit minimalem Schädelhirntrauma und einer Kontrollgruppe im narrativen Diskurs sowie beim Benennen. Dabei zeigten sich Wortfindungsdefizite jedoch im stärkeren Ausmaß in der Benennaufgabe (sowohl für die Normsprecher als auch für die Schädelhirntraumapatienten) und die Autoren argumentieren, dass Aufgaben zum konfrontativen Benennen vergleichsweise sensitiver für die Identifizierung subtiler Defizite bei minimalem Schädelhirntrauma sind. Dies könnte ein Indiz dafür sein, dass Leistungen in relativ umschriebenen Aufgaben besser zur Identifizierung eines Leistungsdefizits geeignet sind als vergleichsweise komplexe Aufgaben wie etwa der narrative Diskurs oder die Spontansprache. Der Wortabruf im narrativen Diskurs wurde in der Studie von King und Kollegen (2006a) anhand von Bildvorlagen elizitiert und beruhte daher auf einem stark strukturierten Thema. Es ist denkbar, dass Analysen der Spontansprache, die auf semi-standardisierten Interviews beruhen, zu anderen Ergebnissen führen. Dies ist in zukünftigen Studien zu untersuchen. Eine Analyse von Spontansprache- und Wortgenerierungscharakteristika bei Multipler Sklerose ist in Vorbereitung (Thiele et al., in Vorbereitung).

Evaluation der Verbgenerierungsleistungen

Darüber hinaus ergeben sich weitere Implikationen aus der Evaluation der Verbgenerierungsleistungen in Studie 1 (vgl. Kapitel 3). Insbesondere Patienten mit chronisch-progredienter Multipler Sklerose zeigten schlechtere Leistungen bei der Verbgenerierung im Vergleich zu Patienten mit schubförmig-remittierendem Verlauf der Erkrankung und Sprachgesunden. Leistungen in der Verbgenerierung werden vor allem im Zusammenhang mit fronto-subkortikalen Schaltkreisen beschrieben (Davis et al., 2010; Östberg et al., 2005; Piatt et al., 1999a). Piatt und Kollegen (1999a) konnten Patienten mit Parkinson und Demenz von Patienten mit Parkinson ohne Demenz anhand der Leistungen in einer Verbgenerierungsaufgabe voneinander dissoziieren und beschreiben einen Zusammenhang zwischen Verbgenerierungsperformanz und fronto-striataler Pathologie. Darüber hinaus weisen weitere Befunde in diese Richtung. So konnten beispielsweise Östberg und Kollegen (2005) in einer Untersuchung von Patienten mit subjektiv empfundenen kognitiven Beeinträchtigungen und Patienten mit *mild cognitive impairment* sowie von Patienten mit Demenz vom Alzheimer Typ eine Minderung in der Verbgenerierungsleistung als sensitives Maß zur Identifizierung einer beginnenden demenziellen Symptomatik im Sinne eines *mild cognitive impairment* identifizieren.

Patienten mit chronisch-progredienter Multipler Sklerose produzierten bei der Verbgenerierung signifikant weniger korrekte Nennungen im Vergleich zu den anderen beiden untersuchten Gruppen (vgl. Kapitel 3). Zu-

dem zeigten Patienten mit chronisch-progredienter Multipler Sklerose auch im DemTect (Kalbe et al., 2004), einem Screening für leichte kognitive Beeinträchtigungen bzw. frühe Symptome einer Demenz, signifikant schlechtere Leistungen im Vergleich zu den Sprachgesunden und den Patienten mit schubförmig-remittierender Multipler Sklerose. Zwei Drittel der untersuchten Patienten mit chronisch-progredienter Multipler Sklerose zeigten im DemTect Hinweise auf leichte kognitive Beeinträchtigungen. Zusammengefasst und in Übereinstimmung mit den Befunden von Piatt und Kollegen (1999a) sowie von Östberg et al. (2005) weisen die Ergebnisse darauf hin, dass defizitäre Verbgenerierungsleistungen auch bei Multipler Sklerose ein Indiz für die Entwicklung einer demenziellen Symptomatik sein können.

Vor dem Hintergrund der Fragestellungen dieser Dissertation lässt sich an dieser Stelle ein weiteres Fazit ziehen. Die Evaluation von Leistungen in semantischen Wortgenerierungsaufgaben scheint bei unterschiedlichsten Störungen verschiedener Ätiologien und den für sie typischen Läsionslokalisationen sinnvoll einsetzbar zu sein um Leistungsdefizite zu identifizieren. Leistungen bei der Verbgenerierung scheinen ein verlässlicher Indikator für eine beginnende demenzielle Symptomatik zu sein.

Der folgende Abschnitt befasst sich mit der Beschreibung eines Einzelfalls. Auch hier steht die Evaluation der Verbgenerierungsleistung im Vordergrund.

6.2.3 Sozial-kognitive Charakteristika von Wortgenerierungsleistungen: Ein Einzelfall

Die Betrachtung eines Einzelfalls aus der Studie zur Evaluation der Wortgenerierung bei Patienten mit Multipler Sklerose (vgl. Kapitel 3) ist insofern interessant, als dieser in besonderem Maße verdeutlicht, dass Auffälligkeiten in einer umschriebenen Wortgenerierungsaufgabe sich auch in sprachlich-kommunikativem Verhalten widerspiegeln können und vice versa.

Im Mittelpunkt der Betrachtung steht der Patient 01MS_CP. Dabei handelt es sich um einen männlichen Versuchsteilnehmer im Alter von 51 Jahren, bei dem zum Zeitpunkt der Testung seit 69 Monaten die Diagnose einer chronisch-progredienten Multiplen Sklerose bestand. Darüber hinaus wiesen die Ergebnisse aus dem DemTect (Kalbe et al., 2004) mit einem Testwert von 12 auf das Bestehen leichter kognitiver Defizite hin. Auf Basis der Diskussion im vorausgehenden Abschnitt (vgl. Abschnitt 6.2.2) bestanden in Bezug auf die Verbgenerierungsleistung bei diesem Patienten bestimmte Erwartungen: Aufgrund der Verlaufsform der Erkrankung sowie der Hinweise auf leichte kognitive Beeinträchtigungen, waren auch reduzierte Leistungen bei der Verbgenerierung zu erwarten. Spezifische Erwartungen bezüglich der

spontansprachlichen Leistungen des Patienten bestanden nicht. Im Folgenden werden nun Charakteristika der Spontansprache des Patienten sowie seine Leistungen bei der Verbgenerierung beschrieben.

Bereits im ersten Kontakt sowie auch in der Spontansprache fiel der Patient 01MS_CP durch ein leicht anzügliches und distanzloses sowie tendenziell unhöfliches und sozial unpassendes Verhalten und dementsprechende Äußerungen auf. Darüber hinaus war ein eingeschränktes Störungsbewusstsein erkennbar und der Patient strahlte eine gewisse Gleichgültigkeit in Bezug auf die folgende Testung aus. Interessanterweise spiegelte sich dies auch in seinen Nennungen bei der Aufgabe zur Verbgenerierung wider. Zunächst sollen zwei Beispiele (siehe Tabelle 6.1 und Tabelle 6.2) aus der Spontansprache des Patienten das beschriebene Verhalten veranschaulichen:

Tabelle 6.1: *Beispiel 1 aus der Spontansprache eines Patienten mit chronisch-progredienter Multipler Sklerose; Thema: Hobby; Kontext: Malen bzw. Kunst als eine der wichtigsten Freizeitbeschäftigungen des Patienten*

- Untersucher:** Gibt es Künstler, die sie besonders mögen?
01MS_CP: ja na es gibt die blauen Reiter zum Beispiel
Untersucher: ja
01MS_CP: es gibt ähm Edvard Munch, fällt mir grad ein
der Schrei ist so ein klassisches Bild kennen sie das?
Untersucher: ja
01MS_CP: ah, kennt ja jeder wahrscheinlich, ähm

Im ersten Beispiel (vgl. Tabelle 6.1) ist insbesondere die letzte Äußerung des Patienten, "kennt ja jeder wahrscheinlich", auffällig. Sie wirkt relativ abwertend und unhöflich. Dieser Eindruck wurde durch die Prosodie weiter verstärkt. Spontansprachliche Äußerungen dieser Art waren in der Spontansprache von 01MS_CP häufig zu finden. Darüber hinaus war im Verlauf zu beobachten, dass die Verantwortung für den erfolgreichen Verlauf des Gesprächs vor allem beim Untersucher lag. Der Patient gab sich zum Teil wortkarg und antwortete erst nach längeren Pausen. Der Gesprächsfluss wurde dann vor allem durch den Untersucher aufrechterhalten.

An anderen Stellen waren die Äußerungen des Patienten demgegenüber als vergleichsweise elaboriert und ausschweifend zu bezeichnen, wie etwa das zweite Beispiel aus der Spontansprache des Patienten zeigt (vgl. Tabelle 6.2). Zu beobachten war zudem, dass die Antworten des Patienten nicht immer kohärent waren und seine Äußerungen oftmals keine Antwort auf die gestellte Frage beinhalteten. Darüber hinaus waren semantische Paraphrasen und Wortfindungsstörungen erkennbar. Die Eindrücke aus dem Gespräch und das Verhalten des Patienten im Verlauf der Testung führten dazu, dass

Kommunikation und Gesprächsverhalten des Patienten als auffällig im Sinne einer kommunikativ-kognitiven bzw. pragmatischen Beeinträchtigung bewertet wurden (Turkstra et al., 1995).

Tabelle 6.2: *Beispiel 2 aus der Spontansprache eines Patienten mit chronisch-progredienter Multipler Sklerose; Thema: Arbeitsalltag; Kontext: Welche Anstrengungen werden für die Akquise von Finanzmitteln unternommen?*

- Untersucher:** [...] Und wie gehen sie da so vor? Was machen sie da so? Gibt es bestimmte Aktionen?
- 01MS_CP:** naja ich sagte ja schon ... ähm so das Motto mein Lebensmotto ist an sich der ja konziliare Prozess ... Frieden, Gerechtigkeit, Bewahrung der Schöpfung. das ist auch das Motte äh in dieser Kinder- und Jugendarbeit ... wir haben ähm einen sehr . ja gesellschaft-, jugendpolitischen Einsatz ... das heißt ähm ... ja wir . positionieren uns in Bezug auf die Programmatik der Kinder und Jugendlichen . und ehm positionieren heißt hier .. wir beziehen Stellung . also zum Beispiel gegen Rassismus . ähm ja .. für . die Umwelt für einen einen einen ja verträglichen Umgang mit der Natur [...]

Bei der Überprüfung der Wortgenerierungsleistungen waren mit einem Fehleranteil von 50% insbesondere die Leistungen bei der Verbgenerierung auffällig. Dies deckt sich mit der vorab auf Basis der Verlaufsform und der Ergebnisse aus dem DemTect formulierten Erwartung. Eine Übersicht über die quantitativen Leistungen in allen fünf Wortgenerierungsaufgaben liefert Tabelle 6.3. Bei der Gegenüberstellung der Ergebnisse aus den verschiedenen Wortgenerierungsaufgaben ist zu erkennen, dass 01MS_CP bei der Verbgenerierung vergleichsweise schlechtere Leistungen zeigte. Die Leistungen in dieser Aufgabe werden daher im Folgenden eingehender betrachtet.

Bei der Elizitierung der Verbgenerierung wurde 01MS_CP, wie alle anderen Versuchsteilnehmer auch, dazu aufgefordert so viele verschiedene "Dinge" wie möglich zu nennen, die ein Lebewesen tun kann. Die von ihm produzierten Äußerungen unterschieden sich jedoch auf den ersten Blick von denen anderer Versuchsteilnehmer; von Sprachgesunden genauso wie von anderen Patienten mit Multipler Sklerose. Auch sein Verhalten bei Durchführung der Aufgabe war als auffällig zu beurteilen. Zur Veranschaulichung findet sich im Folgenden eine Auflistung der Nennungen (siehe Tabelle 6.4).

6.2 Evaluation von Wortgenerierungsleistungen

Tabelle 6.3: Leistungen des Patienten 01MS_CP in den simplen und komplexen lexikalischen und semantischen Wortgenerierungsaufgaben sowie bei der Verbgenerierung; N. gesamt = Anzahl Nennungen gesamt; N. korrekt = Anzahl Nennungen korrekt; FA(%) = Fehleranteil in Prozent

Aufgabe	N. gesamt	N. korrekt	FA(%)
LEX_M	12	11	8.3 %
SEM_Tiere	25	24	4.0 %
LEX_H/T	13	9	30.8 %
SEM_Kleidung/Blumen	15	12	20.0 %
Verbgenerierung	14	7	50 %

Tabelle 6.4: Beispiel aus der Verbgenerierung eines Patienten mit chronisch-progredienter Multipler Sklerose

Patient: lieben, hassen, fröhlich sein, essen, weinen, ängstlich sein, verdauen, pupsen, Sex haben, sich vereinigen beziehungsweise sich vervielfältigen in dem sie (*Patient grinst*), antizipieren, Perspektive entwickeln, sich korrigieren, politisch denken

Bei der Betrachtung der einzelnen Nennungen bei der Verbgenerierung zeigte sich, dass die Nennungen von 01MS_CP im Vergleich zu den Nennungen anderer Versuchsteilnehmer "anders" waren. Von den insgesamt 14 Nennungen produzierte 01MS_CP insgesamt acht, die von keinem anderen Versuchsteilnehmer bei der Verbgenerierung genannt wurde. Zum Vergleich: andere Versuchsteilnehmer produzierten häufig Nennungen wie etwa: *essen, laufen, schlafen, schmecken, sprechen, springen, trinken, lachen ...*³⁶. Neben emotional gefärbten Wörtern (*lieben, hassen, ...*) produzierte 01MS_CP Nennungen aus dem Verdauungs- und Sexualbereich (*pupsen, Sex haben, ...*), welche von einem anzüglichen Grinsen begleitet waren. Regelbrüche waren häufig zu beobachten und äußerten sich in Form von Mehrwortäußerungen, Umschreibungen oder Wörtern der falschen Wortklasse. Insgesamt konnten nur sieben Nennungen als regelkonform und somit als korrekt gewertet werden. Zu beobachten war, dass die Zahl der Regelbrüche im Verlauf der Bearbeitung der Aufgabe zunahm.

Die Beispiele aus Spontansprache und Wortgenerierung zeigten ein ähnliches Bild. Dabei waren nicht nur die Leistungen an sich auffällig, sondern auch das dabei zu beobachtende Verhalten des Patienten. Neben inkohärenten Äußerungen in der Spontansprache und den analog dazu beobachtbaren Regelbrüchen bei der Verbgenerierung, zeigte sich in beiden Aufgaben ein

³⁶Diese Aufzählung beinhaltet die von allen Versuchsteilnehmern aus Studie 1 (vgl. Kapitel 3) am häufigsten produzierten Verben. Die Nennungen sind nach ihrer Auftrenshäufigkeit sortiert. Alle aufgeführten Verben wurden in Studie 1 bei der Elizitierung der Verbgenerierungsleistungen von mindestens einem Viertel aller Versuchsteilnehmer genannt.

anzüglisches sowie sozial unangemessenes Verhalten. Diese Beobachtungen legen nahe, in einer größeren Stichprobe den Zusammenhang von Charakteristika in der Spontansprache und Leistungen in der Wortgenerierung genauer zu untersuchen (Thiele et al., in Vorbereitung). Dabei können sowohl traditionelle quantitative und qualitative Analysemethoden, als auch die in Studie 3 (vgl. Kapitel 5) beschriebene Methodologie Anwendung finden.

Das hier beschriebene Patientenbeispiel betont die Relevanz einer gleichzeitigen Berücksichtigung sowohl kognitiver als auch sozialer, emotionaler und aktivitätsbezogener Ebenen von Exekutivfunktionen (vgl. Drechsler, 2007). Das von 01MS_CP gezeigte Verhalten sowie die beobachtbaren Äußerungen und Nennungen sind daher auch in Zusammenhang mit der von Drechsler (2007) beschriebenen Taxonomie exekutiver Funktionen interessant (vgl. Abschnitt 2.2). Neben basalen und komplexen kognitiven Aspekten von Exekutivfunktionen weist die Autorin auf die Bedeutung der Aktivitätsregulation sowie der emotionalen und sozialen Regulation hin. Insbesondere die von Drechsler (2007) beschriebenen Charakteristika von Beeinträchtigungen in der sozialen sowie der Aktivitätsregulation finden sich auch bei 01MS_CP wieder. Hinweise auf Einschränkungen in der Aktivitätsregulation zeigen sich bei dem Patienten beispielsweise in Form abweichender, inkohärenter Äußerungen oder der zunehmenden Regelbrüche bei der Verbgenerierung. Defizite der sozialen Regulation äußern sich als anzüglisches, distanzloses und unhöfliches Verhalten und dementsprechende Äußerungen sowie im verwendeten sprachlichen Register.

Als geeignetes Aufgabenparadigma zur Erfassung von Leistungen auf Ebene der Aktivitätsregulation beschreibt Drechsler (2007) unter anderem Flüssigkeitsaufgaben. Die Autorin geht dabei jedoch nicht weiter auf die Art und Weise der Evaluation der Leistungen ein. Der beschriebene Einzelfall zeigt, dass die Evaluation einer Wortgenerierungsaufgabe, als spezifische Form einer Flüssigkeitsaufgabe, neben Defiziten in lexikalisch-semantischer oder kognitiv-exekutiver Verarbeitung auch Einschränkungen auf anderen Ebenen anzeigen kann. Die Evaluation von Wort- bzw. Verbgenerierungsleistungen scheint daher eine geeignete Methode zur Erfassung spezifischer Leistungsdefizite.

Vor dem Hintergrund der Fragestellungen dieser Dissertation lässt sich aus diesem Einzelfall ein weiteres Fazit ziehen: Leistungen in einer relativ umschriebenen Verbgenerierungsaufgabe können sich in der Spontansprache widerspiegeln und vice versa. Bei der Evaluation von spezifischen Wortgenerierungsleistungen kann neben quantitativen und qualitativen Aspekten der Analyse auch die Beobachtung des Verhaltens bei der Durchführung der Aufgabe zur Diagnosestellung beitragen.

Im nun folgenden Abschnitt wird der Einsatz von *machine learning*-

Techniken bei der Evaluation von Wortgenerierungsleistungen diskutiert.

6.2.4 Evaluation von Wortgenerierungsleistungen mit *machine learning*-Techniken

Neben traditionellen Analysemethoden werden vermehrt *machine learning*-Techniken zur Identifizierung spezifischer Störungsbilder eingesetzt (Costafreda et al., 2011; Fraser et al., 2012; Gaspers et al., 2012; Järvelin, 2008; Järvelin & Juhola, 2011; Maier et al., 2011; Lehr et al., 2012; Prud'hommeaux & Roark, 2011; Prud'hommeaux et al., 2011; Roark et al., 2007, 2011; Wilson et al., 2009). Die Anwendung von *machine learning*-Techniken als Methode zur Unterstützung der Diagnosestellung ist Teil der sogenannten Gesundheitsinformatik (*health informatics*). Diese stellt eine wachsende Disziplin dar, die sich mit den Anforderungen einer sich immer mehr technologisierenden Welt auseinandersetzt (Hersh, 2009). Nach Hersh (2009) befasst sich die Gesundheitsinformatik "with the optimal use of information, often aided by the use of technology, to improve individual health, health care, public health, and biomedical research" (Hersh, 2009, S. 2). Im Mittelpunkt steht demnach die Nutzung verschiedenster Technologien mit dem Ziel der Verbesserung der Gesundheit, sowohl bezogen auf Individuen als auch auf das öffentliche Gesundheitssystem.

Auch in der vorliegenden Dissertation wurden *machine learning*-Techniken als Methodologie zur Identifizierung kommunikativ-kognitiver Defizite unterschiedlicher Genese eingesetzt. Beim *machine learning* werden Klassifikatoren auf einem Trainingsdatenset trainiert und lernen anhand dieses Datensets und der spezifischen Charakteristika der Trainingsdaten Kategoriezugehörigkeiten. Mit dem Ziel, neue Instanzen den gelernten Kategorien zuzuordnen, wird das resultierende Modell auf sogenannten Testdaten angewendet (Bishop, 2006; Marsland, 2009). Auf Basis von Mustererkennung ist durch den Einsatz von *machine learning*-Techniken eine reliable Identifizierung und Abgrenzung möglich, was durch verschiedene Studien bestätigt wurde (Costafreda et al., 2011; Fraser et al., 2012; Järvelin, 2008; Järvelin & Juhola, 2011; Maier et al., 2011; Lehr et al., 2012; Quintana et al., 2012; Roark et al., 2011; Wilson et al., 2009). Auch die beiden Studien dieser Dissertation, die sich mit dem Einsatz von *machine learning*-Techniken bei der Evaluation von Wortgenerierungsdaten zur Diagnose verschiedener Sprach- und Kommunikationsstörungen befasst haben, bestätigen die Anwendbarkeit dieser Methode. In Studie 2, die sich mit der Klassifikation von Aphasikern, Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden beschäftigte, zeigte die Evaluation der Wortgenerierungsdaten mit *machine learning*-Techniken, dass sich Aphasiker anhand der *Anzahl korrekter Nennungen* reliabel klassifizieren ließen. Dies ist ein Ergebnis, das sich mit Befunden aus der Literatur deckt und auch als anerkannt gilt: Aphasiker

fallen bei der Analyse von Wortgenerierungsleistungen insbesondere durch eine reduzierte *Anzahl korrekter Nennungen* auf (Basso et al., 1997). Die Studien 2 und 3 konnten zudem zeigen, dass eine Evaluation der *Anzahl korrekter Nennungen* mit *machine learning*-Techniken nicht zu einer reliablen Dissoziation zwischen Patienten mit kommunikativ-kognitiven Defiziten und Sprachgesunden führt. Zeitbasierte Auswertungsparameter schienen besser für die Identifizierung kommunikativ-kognitiver Defizite geeignet. Auch dieses Ergebnis ist in Einklang mit Befunden anderer Studien, die beschreiben, dass traditionelle Maße, die die Korrektheit einer Leistung abbilden, weniger gut zur Diagnose vergleichsweise geringer kommunikativ-kognitiver Beeinträchtigungen geeignet sind (Barrow et al., 2003, 2006; King et al., 2006a,b).

Dass sich mit *machine learning*-Techniken klinisch etablierte Befunde (wie etwa die reduzierte *Anzahl korrekter Nennungen* im Falle der Aphasie) bestätigen lassen, kann als Indiz für die Anwendbarkeit dieser Methodologie bei der Evaluation von Wortgenerierungsleistungen gesehen werden. Der Einsatz von *machine learning*-Techniken scheint demnach eine valide Methode zur Diagnose sprachlicher und kommunikativ-kognitiver Defizite auf Basis von Wortgenerierungsdaten zu sein.

Die Analysen aus Studie 3 zeigten auch, dass eine Meta-Klassifikation, das heißt das Zusammenführen individueller Klassifikationsergebnisse zu einer Gesamtentscheidung anhand eines übergeordneten Klassifikators, das Klassifikationsergebnis weiter verbessern kann. Mit Einsatz des Meta-Klassifikators konnten nahezu alle Versuchsteilnehmer der Studie korrekt klassifiziert werden, was im Vergleich zur relativ simplen Klassifikation auf Basis von Einzelvariablen einer erheblichen Verbesserung entspricht. Bei der Klassifikation auf Basis von Einzelvariablen konnten maximal Dreiviertel aller Versuchsteilnehmer korrekt identifiziert werden (vgl. Experiment 1b in Kapitel 5). Das Ergebnis einer besseren Klassifikation bei Verwendung eines Meta-Klassifikators bestätigt die von Jain (2000) beschriebene Annahme einer verbesserten Identifizierung von Kategoriezugehörigkeiten durch die Kombination individueller Entscheidungen (vgl. auch Abschnitt 2.6).

Darüber hinaus weisen verschiedenen Studien, die sich mit der Anwendung von *machine learning*-Techniken bei verschiedenen Störungsbildern befasst haben, darauf hin, dass das Einbeziehen zusätzlicher kognitiver, sprachlicher und/oder kommunikativer Leistungen die Dissoziation zwischen verschiedenen Patientengruppen und Sprachgesunden weiter fazilitiert (vgl. Quintana et al., 2012; Wilson et al., 2009). Einen ersten Hinweis darauf geben auch die Ergebnisse der Meta-Klassifikation in Studie 3. Den verschiedenen Wortgenerierungsaufgaben liegen sich zum Teil überlappende aber auch divergente Verarbeitungsprozesse zugrunde (Basso et al., 1997; Nutter-Upham et al., 2008; Troyer et al., 1998). Die Ergebnisse aus den drei Experimenten zeigen, dass die besten Klassifikationsresultate damit erreicht werden können, wenn die Ergebnisse aus den verschiedenen Wortgenerie-

rungsaufgaben miteinander kombiniert werden. Dies kann als Indiz dafür gesehen werden, dass das Zusammenführen von Leistungen in unterschiedlichen Aufgaben, die auf zum Teil divergenten Verarbeitungsprozessen beruhen, die Klassifikation erleichtert.

Neben der Klassifikationsgenauigkeit (*accuracy*) als primäres Evaluationsmaß kann die Güte der Klassifikation anhand weiterer Parameter bestimmt werden, so beispielsweise mit Angaben zu *Sensitivität* oder *Spezifität* (Bortz & Lienert, 2008; vgl. auch Abschnitt 2.6). Die Berücksichtigung weiterer Evaluationsmaße ermöglicht Aussagen über relevante Eigenschaften diagnostischer Verfahren und Prozeduren (vgl. Abschnitt 2.5). In Studie 3 wurden in Experiment 2b (vgl. Abschnitt 5.4.2) bei der Klassifikation mit aufgabenspezifischen Variablenkombinationen zusätzlich zur *accuracy* auch jeweils *Precision* und *Recall* für die beiden untersuchten Gruppen berechnet. Während die *Precision* angibt, mit welcher Wahrscheinlichkeit ein positives Testergebnis die korrekte Diagnose widerspiegelt, macht der *Recall*-Wert als Sensitivitätsmaß Aussagen darüber, mit welcher Wahrscheinlichkeit eine positive Diagnose mit einem Test bestätigt werden kann (Bortz & Lienert, 2008). Durch diese zusätzlichen Evaluationsmaße konnten in Studie 3 Aussagen darüber gemacht werden, wieviel Prozent der untersuchten Patienten mit kommunikativ-kognitiven Defiziten als sprachgesund fehlklassifiziert wurden. Ist der Anteil fehlklassifizierter Patienten bekannt, sind nur bei diesen Patienten weitere diagnostische Prozeduren notwendig, um Aufschluss über das Vorliegen spezifischer Defizite zu erhalten.

Anzumerken ist an dieser Stelle, dass der Einsatz von *machine learning*-Techniken zur Diagnosestellung bislang in erster Linie ein Instrument der Forschung darstellt. Auf Basis großer Datensätze können spezifische Charakteristika oder Muster von Leistungsdefiziten erkannt werden. Die so identifizierten Parameter können anschließend für die klinische Praxis wertvolle Informationen für das diagnostische und therapeutische Vorgehen bei verschiedenen Störungsbildern liefern. Der Einsatz von *machine learning*-Techniken als universelles Diagnostikum in der klinischen Praxis ist derzeit (noch) nicht realisierbar.

Vor dem Hintergrund der Fragestellungen dieser Dissertation lässt sich an dieser Stelle jedoch festhalten, dass die Evaluation von Wortgenerierungsleistungen mit *machine learning*-Techniken eine geeignete Methode zur Identifizierung verschiedener Störungsbilder zu sein scheint.

Es folgen nun ein abschließendes Fazit sowie ein Ausblick auf weitere Forschung zur Evaluation von Wortgenerierungsleistungen.

6.3 Fazit

An dieser Stelle soll rekapituliert werden, welche Bedeutung die in den vorausgehenden Abschnitten beschriebenen Befunde für das diagnostische Vorgehen haben. Dabei soll auch auf die in Kapitel 2 dargestellten Aspekte einer Diagnostik eingegangen werden.

Nicht alle Störungsbilder wurden durch die Evaluation der *Anzahl korrekter Nennungen* identifiziert. Insbesondere Patienten mit kommunikativ-kognitiven Defiziten konnten mit anderen Variablen besser von Sprachgesunden (und Patienten mit Aphasie) dissoziiert werden. Dieses Ergebnis zeigt, dass es für die Diagnose vergleichsweise gering ausgeprägter Beeinträchtigungen erforderlich sein kann, bestehende Verfahren und Analysemethoden zu erweitern. Um auch bei subtilen Defiziten eine verlässliche Diagnosestellung zu gewährleisten, sollten insbesondere Analysen zeitlicher Parameter in Evaluationsprozeduren aufgenommen werden. Auch Ergebnisse anderer Studien weisen darauf hin, dass die Analyse von Normwerten und anderen quantitativen Maßen, die in erster Linie die Korrektheit der gezeigten Leistungen widerspiegeln, nicht immer ausreichend zu sein scheinen (Barrow et al., 2003, 2006; King et al., 2006b; Lezak et al., 2004).

Für die Diagnose kommunikativ-kognitiver Defizite werden vor allem nicht-standardisierte Verfahren als geeignet beschrieben, da sich bei diesen Patienten Leistungsbeeinträchtigungen vor allem in unstrukturierten Settings zeigen sollen (Coelho et al., 2005). Die beschriebenen Studien, die sich mit der Klassifikation von Patienten mit kommunikativ-kognitiven Defiziten befassen haben (Studie 2 und Studie 3) zeigen jedoch, dass auch mit standardisierten Verfahren (Regensburger Wortflüssigkeitstest, Aschenbrenner et al., 2000) Leistungsunterschiede aufgedeckt werden können. Dies jedoch nur, wenn anstelle einer quantitativen Analyse, qualitative Auswertungsparameter einbezogen werden. Demnach scheint nicht die Wahl eines standardisierten versus nicht-standardisierten Verfahrens an sich ausschlaggebend für die Identifizierung eines Leistungsdefizits bei kommunikativ-kognitiven Defiziten zu sein. Vielmehr scheint die Wahl einer quantitativen oder qualitativen Analyse eine entscheidende Rolle zu spielen.

Die beschriebenen Analysen haben gezeigt, dass insbesondere semantische Aufgaben gut zwischen spezifischen Störungsbildern dissoziieren und daher bei den unterschiedlichsten Störungen verschiedener Ätiologien sinnvoll einsetzbar sein sollten, um Leistungsdefizite zu identifizieren. Dennoch sollten bei einer individuellen Diagnostik Leistungen in verschiedenen Wortgenerierungsaufgaben erhoben werden. So konnten beispielsweise in Studie 1 Leistungsunterschiede zwischen Patienten mit Multipler Sklerose und Sprachgesunden auch für die lexikalische Wechsel-Aufgabe gefunden werden. In Studie 2 zeigten sich die besten Ergebnisse für die Klassifikati-

on zwischen Aphasikern und Sprachgesunden ebenfalls für die lexikalische Wechsel-Aufgabe. Auch Nutter-Upham und Kollegen (2008) betonen die Relevanz verschiedener Wortgenerierungsaufgaben zur Identifizierung spezifischer Störungen bei der Diagnosestellung. Darüber hinaus liefert die Evaluation von Verbgenerierungsleistungen (*action fluency*) eine weitere Möglichkeit zur Diagnose von Leistungsdefiziten. Insbesondere zur Identifizierung beginnender demenzieller Symptome scheinen Verbgenerierungsleistungen geeignet zu sein (Davis et al., 2010; Östberg et al., 2005; Piatt et al., 1999a). Hinweise darauf konnten auch in Studie 1 bei der Analyse von Verbgenerierungsleistungen bei Multipler Sklerose gefunden werden.

Die Überprüfung von Verbgenerierungsleistungen erfolgt zurzeit nicht-standardisiert. Das heißt zum einen, dass die zur Elizitierung verwendeten Instruktionen sich unterscheiden können, was eine Vergleichbarkeit der Leistungen zwischen verschiedenen Studien aber auch auf Patientenebene im Rahmen einer individuellen Diagnostik erschwert (Lezak et al., 2004). Zum anderen besteht auch kein Konsens bezüglich der quantitativen Auswertung von Verbgenerierungsdaten. Während nach Pekkala (2012) auch Verben mit Argumentstruktur als korrekte Nennung gezählt werden (z.B. *Tennis spielen, Teig kneten*), berücksichtigen andere Autoren (z.B. Piatt et al., 1999a) lediglich Einzelwortäußerungen (z.B. *spielen, laufen, etc...*) als korrekte Nennungen. Bezüglich der Instruktionen und Auswertung herrscht daher weiterer Klärungsbedarf und eine standardisierte Herangehensweise wäre wünschenswert.

Wie in Kapitel 2 beschrieben, beruht eine Befunderhebung bzw. Diagnosestellung im Wesentlichen auf Beobachtung. Das heißt, zum einen auf der Beobachtung quantitativer sowie qualitativer Leistungen in einer Aufgabe selbst, zum anderen aber auch auf der Beobachtung des Verhaltens bei der Ausführung einer Aufgabe (Kaplan, 1988; Lezak, 2004; (vgl. Kapitel 2, Abschnitt 2.5). Die Wichtigkeit dieser beiden Aspekte kann mit den Ergebnissen aus den drei Studien und dem beschriebenen Einzelfall weiter betont werden. Insbesondere der in Abschnitt 6.2.3 beschriebene Einzelfall hat gezeigt, dass sich auch in einer relativ umschriebenen Wortgenerierungsaufgabe sozial-kognitive Auffälligkeiten zeigen können, die das Verhalten in der sozialen Interaktion in gewisser Weise widerspiegeln. Eine fundierte Diagnostik sollte daher immer beide Komponenten der Beobachtung bei der Befunderhebung berücksichtigen. Auch hier zeigt sich wieder, wenn auch auf einer anderen Ebene, dass die gleichzeitige Betrachtung quantitativer und qualitativer Aspekte einer Leistung die Diagnosestellung faszilitieren kann (Kaplan, 1988; Lezak et al., 2004).

Die Ergebnisse der beschriebenen Studien unterliegen gewissen Einschränkungen, welche sich auf die Generalisierbarkeit der Befunde auswirken. Wortgenerierungsleistungen wurden in drei bzw. vier unterschiedlichen klinischen

Populationen untersucht. Bei Patienten mit Aphasie, Patienten mit kommunikativ-kognitiven Defiziten sowie bei Patienten mit Multipler Sklerose unterschiedlicher Verlaufsform. Für die Diagnose von Aphasien stehen standardisierte Verfahren zur Diagnose zur Verfügung (z.B. Huber et al., 1983). Aufgrund eines Mangels an standardisierten Diagnosekriterien, wurden kommunikativ-kognitive Defizite auf Basis von Expertenurteilen in Kombination mit einem Ausschluss sprachsystematischer Defizite anhand des AAT-Screenings (Huber et al., 1983) diagnostiziert. Patienten mit Multipler Sklerose wurden auf Basis der ärztlichen Diagnose und der jeweiligen Verlaufsform in die Studie aufgenommen. Es ist daher anzunehmen, dass die Gruppe der Aphasiepatienten relativ zur Gruppe der Patienten mit kommunikativ-kognitiven Defiziten und den beiden Gruppen der Multiple Sklerose Patienten homogener in Bezug auf ihre Leistungen waren, was eine einfacherer Dissoziation bzw. bessere Klassifikationsergebnisse begründen könnte. Als weiterer Punkt ist die relativ geringe Stichprobengrößen in den drei Studien anzumerken. Insbesondere die Gruppe der Patienten mit chronisch-progredienter Multipler Sklerose ist unterrepräsentiert. Bevor allgemeinere Aussagen getroffen werden können, ist die Untersuchung einer größeren Stichprobe erforderlich. Die weitere Datenerhebung hierzu wurde bereits begonnen.

Abschließend sollen nun die in Kapitel 2 (siehe Abschnitt 2.7) formulierten Fragen noch einmal aufgegriffen und beantwortet werden:

(a) *Ist die Evaluation von Wortgenerierungsleistungen zur Diagnose kommunikativ-kognitiver Defizite und anderer Sprach- und Kommunikationsstörungen geeignet?*

Die Evaluation von Wortgenerierungsleistungen stellt eine gute Möglichkeit zur Diagnose kommunikativ-kognitiver Defizite sowie weiterer Sprach- und Kommunikationsstörungen dar. In Abhängigkeit vom vorliegenden Störungsbild scheinen jedoch unterschiedliche Methodologien bzw. verschiedene Aspekte quantitativer sowie qualitativer Evaluation für eine sichere Identifizierung einer spezifischen Störung geeignet.

(b) *Welche Auswertungsparameter ermöglichen eine reliable Diagnosestellung? Unterscheiden sich die Auswertungsparameter für die Diagnose verschiedener Störungsbilder?*

Eine verlässliche Diagnosestellung ist sowohl anhand quantitativer als auch auf Basis qualitativer Auswertungsparameter möglich. In Abhängigkeit vom untersuchten Störungsbild kann die Schwerpunktsetzung bei der Evaluation variieren. Für die Diagnose einer Aphasie scheint der Aufwand einer Evaluation zusätzlicher lexikalisch-semantischer, sublexikalischer sowie

zeitlicher Parameter nicht gerechtfertigt, da aphasische Defizite reliabel anhand der *Anzahl korrekter Nennungen* identifiziert werden können. Gleiches gilt für die Identifizierung von Leistungsdefiziten im Rahmen einer Multiplen Sklerose. Auch hier scheinen quantitative Auswertungsparameter, die die Korrektheit einer Leistung widerspiegeln als geeignetes Maß zur Leistungsbeurteilung. Für die Diagnose vergleichsweise geringer kommunikativ-kognitiver Defizite scheint hingegen die Analyse zusätzlicher, insbesondere zeitlicher Parameter, die Identifizierung eines Defizits zu erleichtern. Das heißt, um auch relativ subtile Defizite aufdecken zu können, sollten zusätzlich zum traditionell verwendeten Evaluationsmaß *Anzahl korrekter Nennungen* weitere qualitative Charakteristika bei der Evaluation von Wortgenerierungsleistungen berücksichtigt werden.

(c) *Welche Rückschlüsse können aus reduzierten Wortgenerierungsleistungen oder spezifischen Leistungsunterschieden gezogen werden?*

Insbesondere semantische Wortgenerierungsaufgaben scheinen gut zwischen verschiedenen Populationen zu dissoziieren, was Defizite in der semantischen Verarbeitung nahelegt. Die Evaluation von Wortgenerierungsleistungen in semantischen Aufgaben scheint bei unterschiedlichsten Störungen verschiedener Ätiologien sinnvoll einsetzbar zu sein, um Leistungsdefizite identifizieren zu können. Aber auch Defizite in lexikalischen oder kognitiv-exekutiv anspruchsvolleren Wechsel-Aufgaben lassen sich beobachten und deuten auf Beeinträchtigungen lexikalischer und/oder kognitiv-exekutiver Verarbeitungsprozesse hin. Leistungen bei der Verbgenerierung scheinen ein Indikator für eine beginnende demenzielle Symptomatik zu sein. Darüber hinaus lässt sich erkennen, dass Leistungen, die sich in einer umschriebenen Wortgenerierungsaufgabe manifestieren, sich auch in spontansprachlichen Charakteristika widerspiegeln können und vice versa. Dies ist vor allem für diagnostische Überlegungen von Bedeutung.

Die Berücksichtigung individueller Leistungen in verschiedenen Wortgenerierungsaufgaben scheint eine geeignete Methode zur Identifizierung von Leistungsdefiziten unterschiedlicher Genese zu sein. Wortgenerierungsleistungen sollten daher stets in verschiedenen Aufgaben erhoben und evaluiert werden.

(d) *Welche der verwendeten Methodologien kann für die Evaluation von Wortgenerierungsleistungen sinnvoll eingesetzt werden?*

Neben eher traditionellen Methodologien, das heißt die Anwendung quantitativer und qualitativer Analysen bei der klinischen Evaluation von Wortgenerierungsleistungen, kann der Einsatz von *machine learning*-Techniken als aussichtsreiche Methode zur Identifizierung von Patienten mit spezifischen sprachlichen oder kommunikativ-kognitiven Defiziten und deren Ab-

grenzung gesehen werden. Mit dem Einsatz von *machine learning*-Techniken zur Diagnose spezifischer Störungsbilder anhand kognitiver oder linguistischer Daten scheint sich eine Methode zu entwickeln, die langfristig die klinische Analyse sinnvoll ergänzen kann. Derzeit stellen *machine learning*-Techniken in erster Linie ein Instrument der Forschung dar, welches zur Identifizierung charakteristischer Leistungsprofile und Störungsmuster eingesetzt wird.

6.4 Ausblick

Die beschriebenen Studien und die mit ihnen gewonnenen Erkenntnisse eröffnen neue Fragen, die Gegenstand zukünftiger Forschung sein sollten.

Die Analyse von Leistungen bei der Verbgenerierung zeigte sich als reliable Methode, um Patienten mit chronisch-progredienter Multipler Sklerose von Sprachgesunden sowie von Patienten mit schubförmig-remittierender Multipler Sklerose zu unterscheiden. Interessant wäre nun, Verbgenerierungsleistungen bei Patienten mit kommunikativ-kognitiven Defiziten (z.B. nach Schädelhirntrauma) zu analysieren. Kommunikativ-kognitive Defizite werden unter anderem nach Schädigung des (prä-)frontalen Kortex oder diffusen subkortikalen Läsionen berichtet. Da die Leistungen bei der Verbgenerierung vor allem im Zusammenhang mit fronto-subkortikalen Schaltkreisen gesehen werden (Davis et al., 2010; Östberg et al., 2005; Piatt et al., 1999a), könnte die Evaluation von Verbgenerierungsleistungen bei dieser Population weitere Erkenntnisse liefern und gegebenenfalls zur Identifizierung kommunikativ-kognitiver Defizite beitragen. King und Kollegen (2006b) konnten in einer Studie zum Wortabruf von Nomen und Verben bei Patienten mit leichtem Schädelhirntrauma Hinweise auf Unterschiede in der Leistung bei der Verarbeitung von Verben und Nomen finden. In der Studie wurden verschiedene Aufgaben zum Abruf von Nomen (Bildbenennen, Satzergänzung, Benennen nach Umschreibung etc.) im Vergleich zu einer Aufgabe zum Verbabruf (Bildbenennen) untersucht. Die Analyse der Verbgenerierungsfähigkeit sowie ein Vergleich von Verbgenerierungsleistungen gegenüber Leistungen in einer Aufgabe die das Produzieren von Nomen erfordert (z.B. Kategorie *Tiere*), könnte die bisherigen Befunde sinnvoll ergänzen.

Die Analyse von Verbgenerierungsleistungen bietet darüber hinaus weitere Möglichkeiten einer qualitativen Analyse der Nennungen. Denkbar wären hier z.B. eine detaillierte Analyse der Fehlertypen, die verschiedene Formen der Regelbrüche und Perseverationen berücksichtigt. Aber auch die Evaluation spezifischer *Cluster*, z.B. phonematische (*laufen, lachen, landen...*), semantische (z.B. *gehen, laufen, rennen, ...* oder *sprechen, reden, erzählen, ...*) und/oder Funktionscluster (z.B. Zustands-, Verlaufs- oder Tätigkeitsverben) sowie eine Einteilung der Verben hinsichtlich ihres Emotionsgehalts (z.B.

emotionsbezogen vs. nicht-emotionsbezogen³⁷) könnten Bestandteil weiterer Untersuchungen sein.

Wortgenerierung in semantischen Kategorien wird häufig anhand der Kategorie *Tiere* elizitiert (Schwartz, Baldo, Graves, & Brugger, 2003). Nach Barrow und Kollegen (2003) ist die Wortproduktionen in natürlichen Kategorien (z.B. *Tiere, Blumen, ...*) schwieriger als in artifiziellen Kategorien (z.B. *Werkzeug, Kleidung, ...*). Als natürliche Kategorien gelten solche, deren konstituierende Items nicht vom Menschen geschaffen wurden. Items aus natürlichen Kategorien werden als weniger gut unterscheidbar angesehen, da sie weniger distinkte Merkmale aufweisen sollen. In Aufgaben zum Benennen wird ihnen zudem eine höhere Antwortlatenz zugeschrieben (Barrow et al., 2006). Verschiedene semantische Kategorien werden darüber hinaus mit Unterschieden in der neuronalen Aktivierung in Verbindung gebracht, so auch natürliche gegenüber artifiziellen Kategorien. Den natürlichen Kategorien wird dabei ein größeres Areal neuronaler Aktivierung zugeschrieben, was als Hinweis darauf gewertet wird, dass sie mehr kognitive Kapazität beanspruchen als andere Kategorien (Barrow et al., 2006). Interessant wäre nun, in zukünftigen Studien zur Wortgenerierung sowohl natürliche, als auch artifizielle Kategorien bei der Evaluation von Wortgenerierungsleistungen in unterschiedlichen klinischen Populationen einzubeziehen, um mögliche kategoriespezifische Unterschiede aufdecken zu können. Wenn natürliche Kategorien mehr kognitive Ressourcen erfordern ist anzunehmen, dass sie gegenüber artifiziellen Kategorien sensitiver für die Identifizierung insbesondere kognitiv-exekutiver Leistungsdefizite sein sollten. In diesem Zusammenhang und mit Bezug auf Abschnitt 6.2.2 könnte dann auch die Frage Berücksichtigung finden, ob sich die besserer Dissoziation anhand semantischer Wortgenerierungsaufgaben eventuell auch auf kategoriespezifische Effekte bzw. die damit verbundenen kognitiven Anforderungen zurückführen lässt.

Auch der Einsatz von *machine learning*-Techniken bei der Evaluation von Wortgenerierungsleistungen eröffnet weitere interessante Möglichkeiten. Die Evaluation von Wortgenerierungsleistungen mit *machine learning*-Techniken könnte etwa weitere Klassifikationsaufgaben fokussieren. So bestehen beispielsweise kontroverse Diskussionen dahingehend, dass nicht nur schubförmig-remittierende und chronisch-progrediente Verlaufsformen einer Multiplen Sklerose durch unterschiedliche kognitive Leistungsprofile gekennzeichnet sind, sondern dass sich darüber hinaus auch primär von sekundär chronisch-progredienten Verlaufsformen abgrenzen lassen (Kraus et al., 2005; Messinis et al., 2013). Interessant wäre hier zu untersuchen, ob sich mit *machine learning*-Techniken charakteristische Leistungsprofile erkennen lassen, die eine Abgrenzung der drei Verlaufsformen anhand spe-

³⁷Dies stellt sicherlich eher ein Kontinuum dar als zwei starr definierte Entitäten.

zifischer Parameter ermöglichen.

Darüber hinaus ist denkbar, Klassifikationen mehrerer spezifischer Störungsbilder unter Einbezug verschiedener sprachlicher, kognitiver und/oder kommunikativer Leistungen vorzunehmen (z.B. Aphasie vs. kommunikativ-kognitive Defizite vs. Leistungsdefizite bei Multipler Sklerose vs. demenzielle Symptomatiken). Das Ziel wäre dabei, charakteristische Störungsmuster zu identifizieren, um eine Differentialdiagnose erleichtern zu können. Da im Rahmen einer Multiplen Sklerose neben kommunikativ-kognitive Defiziten und seltener beschriebenen sprachsystematischen Beeinträchtigungen auch Symptome einer subkortikalen Demenz diskutiert werden (Lacour et al., 2004; Rao, 1986; Sepulcre et al., 2011), wäre in diesem Zusammenhang vor allem interessant, ob und inwiefern sich die Defizite bei Patienten mit Multipler Sklerose von denen einer Aphasie, kommunikativ-kognitiver Defizite oder einer Demenz abgrenzen lassen.

Zusammenfassend kann festgestellt werden, dass eine reliable Dissoziation zwischen Patienten mit Sprach- und Kommunikationsstörungen und Sprachgesunden nach wie vor eine Herausforderung bedeutet. Die vorliegende Arbeit trägt jedoch insofern zur Auflösung dieser Herausforderung bei, als gezeigt werden konnte, dass die Evaluation quantitativer sowie qualitativer Charakteristika von Wortgenerierungsleistungen eine geeignete Methode zur Identifizierung sprachlicher und kommunikativ-kognitiver Defizite unterschiedlicher Genese darstellt.

Literaturverzeichnis

- Abwender, D. A., Swan, J. G., Bowerman, J. T., & Connely, S. W. (2001). Qualitative analysis of verbal fluency output: Review and comparison of several scoring methods. *Assessment*, 8(3), 323–336.
- Adamovich, B. B., & Henderson, J. (1992). *Scales of Cognitive Ability for Traumatic Brain Injury (SCATBI)*. Pro-Ed.
- Adamovich, B. L. B., & Hutchinson, T. A. (1994). Differential performance of traumatic brain injury subjects and non-brain-injured peers on cognitive tasks. *Clinical Aphasiology*, 22, 257–264.
- Adams, M. L., Reich, A. R., & Flowers, C. R. (1989). Verbal fluency characteristics of normal and aphasic speakers. *Journal of speech and hearing research*, 32(4), 871–879.
- Aha, D. W., & Kibler, D. (1991). Instance-based learning algorithms. *Machine Learning*, 6, 37–66.
- Aichert, I., Marquardt, C., & Ziegler, W. (unveröff.). *Computergestützte Generierung und Analyse von Sprachmaterial nach Struktur- und Frequenzmerkmalen*. München: EKN.
- Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review*, 16(1), 17–42.
- American Speech-Language-Hearing Association (2005). Roles of speech-language pathologists in the identification, diagnosis, and treatment of individuals with cognitivecommunication disorders: Position Statement. URL www.asha.org/policy
- Araujo, N. B. d., Barca, M. L., Engedal, K., Coutinho, E. S. F., Deslandes, A. C., & Laks, J. (2011). Verbal fluency in Alzheimer’s disease, Parkinson’s disease, and major depression. *Clinics (São Paulo, Brazil)*, 66(4), 623–627.
- Arnott, W. L., Jordan, F. M., Murdoch, B. E., & Lethlean, J. B. (1997). Narrative discourse in multiple sclerosis: An investigation of conceptual structure. *Aphasiology*, 11(10), 969–991.

LITERATURVERZEICHNIS

- Arrondo, G., Sepulcre, J., Duque, B., Toledo, J., & Villoslada, P. (2010). Narrative speech is impaired in multiple sclerosis. *European Neurological Journal*, 2(1), 11–40.
- Aschenbrenner, S., Tucha, O., & Lange, K. W. (2000). *RWT – Regensburger Wortflüssigkeits-Test: Handanweisung*. Göttingen: Hofgrefe.
- Azambuja, M. J., Haddad, M. S., Radanovic, M., Barbaso, E. R., & Mansur, L. L. (2007). Semantic, phonologic, and verb fluency in Huntington's disease. *Dementia & Neuropsychologia*, 1(4), 381–385.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423.
- Baddeley, A. (2003a). Working memory and language: An overview. *Journal of Communication Disorders*, 36, 189–208.
- Baddeley, A. (2003b). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829–839.
- Baldo, J. V., Schwartz, S., Wilkins, D. P., & Dronkers, N. (2010). Double dissociation of letter and category fluency following left frontal and temporal lobe lesions. *Aphasiology*, 24(12), 1593–1604.
- Baldo, J. V., Shimamura, A. P., Delis, D. C., Kramer, J., & Kaplan, E. (2001). Verbal and design fluency in patients with frontal lobe lesions. *Journal of the International Neuropsychological Society : JINS*, 7(5), 586–596.
- Barrow, I. M., Hough, M., Rastatter, M. P., Walker, M., Holbert, D., & Rotondo, M. F. (2003). Can within-category naming identify subtle cognitive deficits in the mild traumatic brain-injured patient? *The Journal of Trauma, Injury, Infection and Critical Care*, 54(5), 888–897.
- Barrow, I. M., Hough, M., Rastatter, M. P., Walker, M., Holbert, D., & Rotondo, M. F. (2006). The effects of mild traumatic brain injury on confrontation naming in adults. *Brain Injury*, 20(8), 845–855.
- Barry, D., Bates, M. E., & Labouvie, E. (2008). FAS and CFL forms of verbal fluency differ in difficulty: A meta-analytic study. *Applied Neuropsychology*, 15(2), 97–106.
- Basso, A., Burgio, F., & Pradoni, P. (1997). Semantic category and initial letter word fluency in left-brain-damaged patients. *European Journal of Neurology*, 4, 544–550.
- Beatty, W. W. (2002). Fluency in multiple sclerosis: Which measure is best? *Multiple sclerosis (Houndmills, Basingstoke, England)*, 8(3), 261–264.

- Beatty, W. W., Goodkin, D. E., Beatty, P. A., & Monson, N. (1989). Frontal lobe dysfunction and memory impairment in patients with chronic progressive multiple sclerosis. *Brain and Cognition*, *11*(1), 73–86.
- Beatty, W. W., Testa, J. A., English, S., & Winn, P. (1997). Influences of clustering and switching on the verbal fluency performance of patients with Alzheimer's disease. *Aging, Neuropsychology and Cognition*, *4*(4), 273–279.
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer New York.
- Bittner, R. M., & Crowe, S. F. (2006). The relationship between naming difficulty and FAS performance following traumatic brain injury. *Brain Injury*, *20*(9), 971–980.
- Blake, M. L., Duffy, J. R., Myers, P. S., & Tompkins, C. A. (2002). Prevalence and patterns of right hemisphere cognitive/communicative deficits: Retrospective data from an inpatient rehabilitation unit. *Aphasiology*, *16*(4-6), 537–547.
- Bortz, J., & Lienert, G. A. (2008). *Kurzgefasste Statistik für die klinische Forschung: Leitfaden für die verteilungsfreie Analyse kleiner Stichproben*. Heidelberg: Springer Medizin Verlag, 3 ed.
- Böwering, T. (2012). *Spontansprache bei schubförmiger Multipler Sklerose*. Unveröffentlichte Masterarbeit: Universität Bielefeld.
- Bowers, L., Huisinigh, R., LoGiudice, C., & Orman, J. (2005). *The WORD Test-2: Adolescent*. LinguiSystems.
- Busch, R. M., McBride, A., Curtiss, G., & Vanderploeg, R. D. (2005). The components of executive functioning in traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *27*(8), 1022–1032.
- Butters, M. A., Goldstein, G., Allen, D. N., & Shemansky, W. J. (1998). Neuropsychological similarities and differences among Huntington's disease, multiple sclerosis, and cortical dementia. *Archives of Clinical Neuropsychology*, *13*(8), 721–735.
- Calabrese, P., Kalbe, E., & Kessler, J. (2004). *MUSIC: Multiple Sklerose Inventar Cognition*. biogen idec.

- Callahan, C. C., Johnstone, B., & Stonnington, H. H. (2001). The assessment and rehabilitation of executive function disorders. *Rehabilitation of neuropsychological disorders*, B Johnstone, HH Stonnington. Edwards Brothers, Lillington, NC, (pp. 87–124).
- Capitani, E., Rosci, C., Saetti, M. C., & Laiacona, M. (2009). Mirror asymmetry of category and letter fluency in traumatic brain injury and Alzheimer's patients. *Neuropsychologia*, *47*, 423–429.
- Chiaravalloti, N. D., & Luca, J. d. (2008). Cognitive impairment in multiple sclerosis. *The Lancet*, *7*(12), 1139–1151.
- Coelho, C., Ylvisaker, M., & Turkstra, L. S. (2005). Nonstandardized assessment approaches for individuals with traumatic brain injuries. *Seminars in Speech and Language*, *26*(4), 223–241.
- Coelho, C. A., & DeRuyter, F. (1996). Treatment efficacy: Cognitive communicative disorders resulting from traumatic brain injury in adults. *Journal of Speech and Hearing Research*, *39*, 5–17.
- Coelho, C. A., DeRuyter, F., Kennedy, M. R., & Stein, M. (2008). Treatment efficacy summary: Cognitive-communication disorders resulting from traumatic brain injury.
URL <http://www.asha.org/uploadedFiles/public/TreatmentEfficacySummaries2008.pdf>
- Constantinidou, F., Wertheimer, J. C., Tsanadis, J., Evans, C., & Paul, D. R. (2012). Assessment of executive functioning in brain injury: Collaboration between speech-language pathology and neuropsychology for an integrative neuropsychological perspective. *Brain Inj*, *26*(13-14), 1549–1563.
- Costafreda, S. G., Fu, C. H. Y., Picchioni, M., Touloupoulou, T., McDonald, C., Kravariti, E., Walshe, M., Prata, D., Murray, R. M., & McGuire, P. K. (2011). Pattern of neural responses to verbal fluency shows diagnostic specificity for schizophrenia and bipolar disorder. *BMC psychiatry*, *11*, 18.
- Crosson, B. (1985). Subcortical functions in language: A working model. *Brain and Language*, *25*(2), 257–292.
- Crosson, B. (1999). Subcortical mechanisms in language: Lexical–semantic mechanisms and the thalamus. *Brain and Cognition*, *40*(2), 414–438.
- Crosson, B. (2012). Thalamic mechanisms in language: A reconsideration based on recent findings and concepts. *Brain and Language*.
- Crowe, S. F. (1992). Dissociation of two frontal lobe syndromes by a test of verbal fluency. *Journal of Clinical and Experimental Neuropsychology*, *14*(2), 327–339.

- Crowe, S. F. (1998). Decrease in performance on the verbal fluency test as a function of time: Evaluation in a young healthy sample. *Journal of Clinical and Experimental Neuropsychology*, *20*(3), 391–401.
- Cummings, J. (2007). Pragmatics and adult language disorders: Past achievements and future directions. *Seminars in Speech and Language*, *28*, 96–110.
- Davis, C., Heidler-Gary, J., Gottesman, R. F., Crinion, J., Newhart, M., Moghekar, A., Soloman, D., Rigamonti, D., Cloutman, L., & Hillis, A. E. (2010). Action versus animal naming fluency in subcortical dementia, frontal dementias, and Alzheimer's disease. *Neurocase*, *16*(3), 259–266.
- Davis, G. A. (2007). *Aphasiology: Disorders and clinical practice*. Boston: Pearson, 2nd edition ed.
- De Bleser, R. (2004). *LEMO-Lexikon modellorientiert. Handbuch*. Elsevier, Urban & Fischer.
- Deerwester, S., Dumais, S., Furnas, G., Landauer, T., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, *41*, 391–407.
- Delbeuck, X., Debachy, B., Pasquier, F., & Moroni, C. (2013). Action and noun fluency testing to distinguish between Alzheimer's disease and dementia with Lewy bodies. *Journal of Clinical and Experimental Neuropsychology*, (pp. 1–10).
- Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science*, *23*(4), 517–542.
- DGN (2009). Akuttherapie des ischämischen Schlaganfalls: Leitlinie der Deutschen Gesellschaft für Neurologie (DGN) und der Deutschen Schlaganfallgesellschaft (DSG) in der Deutschen Gesellschaft für Neurologie.
- Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation*, *10*(7), 1895–1923.
- Douglas, J. M. (2010). Relation of executive functioning to pragmatic outcome following severe traumatic brain injury. *Journal of Speech, Language and Hearing Research*, *53*(2), 365–382.
- Downes, J. J., Sharp, H. M., Costall, B. M., Sagar, H. J., & Howe, J. (1993). Alternating fluency in Parkinson's disease: An evaluation of the attentional control theory of cognitive impairment. *Brain*, *116*(4), 887–902.

- Drake, M. A., Carrá, A., Allegri, R. F., & Luetic, G. (2006). Differential patterns of memory performance in relapsing, remitting and secondary progressive multiple sclerosis. *Neurology India*, 54(4), 370–376.
- Drechsler, R. (2007). Exekutive Funktionen. *Zeitschrift für Neuropsychologie*, 18(3), 233–248.
- Duff, M., & Proctor, K., A. and Haley (2002). Mild traumatic brain injury (MTBI): Assessment and treatment procedures used by speech-language pathologists (SLPs). *Brain Injury*, 16(9), 773–787.
- Ellis, A. W. (1982). Spelling and writing (and reading and speaking). In A. W. Ellis (Ed.) *Normality and Pathology in Cognitive Functions*, (pp. 113–146). London: Academic Press.
- Ellis, A. W., Miller, D., & Sin, G. (1983). Wernicke's aphasia and normal language processing: A case study in cognitive neuropsychology. *Cognition*, 15(1), 111–144.
- Fiehler, R. (1990). *Kommunikation und Emotion: Theoretische und empirische Untersuchungen zur Rolle von Emotionen in der verbalen Interaktion*. De Gruyter.
- Fisseni, H.-J. (2004). *Lehrbuch der psychologischen Diagnostik: Mit Hinweisen zur Intervention*. Göttingen: Hogrefe, 3 ed.
- Fraser, K. C., Meltzer, J. A., Graham, N. L., Leonard, C., Hirst, G., Black, S. E., & Rochon, E. (2012). Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. *Cortex; a journal devoted to the study of the nervous system and behavior*.
- Friend, K. B., Rabin, B. M., Groninger, L., Deluty, R. H., Bever, C., & Grattan, L. (1999). Language functions in patients with multiple sclerosis. *The Clinical neuropsychologist*, 13(1), 78–94.
- Fuster, J. (2002). Frontal lobes and cognitive development. *Journal of Neurocytology*, 31, 373–385.
- Gabrilovich, E., & Markovitch, S. (2007). Computing semantic relatedness using Wikipedia-based explicit semantic analysis. In *Proceedings of the 20th International Joint Conference on Artificial intelligence*.
- Garrett, M. F. (1984). The organization of processing structure for language production: Applications to aphasic speech. In D. Caplan, A. R. Lecours, & A. Smith (Eds.) *Biological Perspectives on Language*, (pp. 172–193). Cambridge/ Massachusetts und London: MIT Press.

- Gaspers, J., Thiele, K., Cimiano, P., Foltz, A., Stenneken, P., & Tscherepanow, M. (2012). An evaluation of measures to dissociate language and communication disorders from healthy controls using machine learning techniques. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, IHI '12*, (pp. 209–218). New York, NY, USA: ACM.
- Geurts, J. J. G., Calabrese, M., Fisher, E., & Rudick, R. A. (2012). Measurement and clinical effect of grey matter pathology in multiple sclerosis. *The Lancet Neurology*, *11*(12), 1082–1092.
- Goodglass, H., & Kaplan, E. (1972). *The assessment of aphasia and related disorders*. Philadelphia: Lea and Febiger.
- Gourovitch, M. L., Goldberg, T. E., & Weinberger, D. R. (1996). Verbal fluency deficits in patients with schizophrenia: Semantic fluency is differentially impaired as compared with phonologic fluency. *Neuropsychology*, *10*(4), 573–577.
- Grossman, M. (1981). A bird is a bird is a bird: Making reference within and without superordinate categories. *Brain and Language*, *12*(2), 313–331.
- Hall, J., Harvey, M., Vo, H., & O'Bryant, S. (2011). Performance on a measure of category fluency in cognitively impaired elderly. *Aging, Neuropsychology, and Cognition*, *18*(3), 353–361.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, *11*(1), 10–18.
- Haykin, S. (1999). *Neural networks. A comprehensive foundation*. Prentice Hall, London, United Kingdom, 2 edition.
- Heidler, M. (2010). Kognitive Dysphasien – Ein Klassifikationsmodell für nichtaphasische zentrale Sprachstörungen. *Neurologie & Rehabilitation*, *16*(5), 217–222.
- Heidler, M. D. (2009). Thalamische Aphasien: Ätiologie und Erscheinungsformen von Sprachverarbeitungsstörungen nach Thalamusläsionen. *Zeitschrift für Neuropsychologie*, *20*(2), 127–139.
- Henry, J. D., & Beatty, W. W. (2006). Verbal fluency deficits in multiple sclerosis. *Neuropsychologia*, *44*, 1166–1174.
- Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology*, *18*(2), 284–295.

LITERATURVERZEICHNIS

- Henry, J. D., & Phillips, L. H. (2006). Covariates of production and perseveration on tests of phonemic, semantic and alternating fluency in normal aging. *Neuropsychology, development, and cognition. Section B, Aging, neuropsychology and cognition*, 13(3-4), 529–551.
- Herrmann, T. (2003). Theorien und Modelle der Sprachproduktion. In G. Rickheit, T. Herrmann, & W. Deutsch (Eds.) *Psycholinguistik: Ein internationales Handbuch*, vol. 24, (pp. 213–228). De Gruyter Mouton.
- Herrmann, T., & Grabowski, J. (1994). Sprechen. Heidelberg: Spektrum.
- Hersh, W. (2009). A stimulus to define informatics and health information technology. *BMC Medical Informatics and Decision Making*, 9(1), 24.
- Hillis, A. E., Wityk, R. J., Barker, P. B., Beauchamp, N. J., Gailloud, P., Murphy, K., Cooper, O., & Metter, E. J. (2002). Subcortical aphasia and neglect in acute stroke: The role of cortical hypoperfusion. *Brain*, 125(5), 1094–1104.
- Huber, W., Poeck, K., & Weniger, D. (2006). Klinisch-neuropsychologische Syndrome und Störungen: Aphasie. In K. Poeck, & W. Hartje (Eds.) *Klinische Neuropsychologie*, (pp. 93–173). Thieme.
- Huber, W., Poeck, K., Weniger, D., & Willmes, K. (1983). *Aachener Aphasie Test*. Hogrefe - Verlag für Psychologie.
- Huber, W., Stachowiak, F.-J., Poeck, K., & Kerschensteiner, M. (1975). Die Wernicke-Aphasie. *Journal of Neurology*, 210(2), 77–97.
- Huber, W., & Ziegler, W. (2000). Störungen von Sprache und Sprechen. In W. Sturm, M. Herrmann, & C. W. Wallesch (Eds.) *Lehrbuch der klinischen Neuropsychologie: Grundlagen, Methoden, Diagnostik, Therapie*, (pp. 462–511). Hogrefe & Huber.
- Huijbregts, S. C., Kalkers, N. F., Sonnevile, L. M. d., Groot, V. d., Reuling, I. E., & Polman, C. H. (2004). Differences in cognitive impairment of relapsing remitting, secondary, and primary progressive MS. *Neurology*, 63(2), 335–339.
- Jaecks, P. (2006). *Restaphasie: Eine empirische Untersuchung von linguistischer Symptomatik, Gesprächsverhalten, Differentialdiagnose und Ursache minimal aphasischer Störungen nach Schlaganfall*. Universität Bielefeld, Dissertation.
- Jaecks, P., Hielscher-Fastabend, M., & Stenneken, P. (2012). Diagnosing residual aphasia using spontaneous speech analysis. *Aphasiology*, 26(7), 953–970.

- Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 4–37.
- Järvelin, A. (2008). Comparison of three neural network classifiers for aphasic and non-aphasic naming data. In *Proceedings of the First International Conference on Health Informatics*.
- Järvelin, A., & Juhola, M. (2011). Comparison of machine learning methods for classifying aphasic and non-aphasic speakers. *Computer methods and programs in biomedicine*, 104(3), 349–357.
- Joanette, Y., & Ansaldo, A. I. (1999). Clinical note: Acquired pragmatic impairments and aphasia. *Brain and Language*, 68(3), 529–534.
- John, G. H., & Langley, P. (1995). Estimating continuous distributions in bayesian classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*.
- Jurado, M. A., Mataro, M., Verger, K., Bartumeus, F., & Junque, C. (2000). Phonemic and semantic fluencies in traumatic brain injury patients with focal frontal lesions. *Brain Injury*, 14(9), 789–795.
- Kalbe, E., Kessler, J., Calabrese, P., Smith, R., Passmore, A. P., Brand, M. a., & Bullock, R. (2004). DemTect: A new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *International Journal of Geriatric Psychiatry*, 19(2), 136–143.
- Kammer, T., & Karnath, H.-O. (2006). Manifestationen von Frontalhirn-schädigungen. In H.-O. Karnath, & P. Thier (Eds.) *Neuropsychologie*, Springer-Lehrbuch, (pp. 489–500). Berlin/Heidelberg: Springer-Verlag.
- Kaplan, E. (1988). The process approach to neuropsychological assessment. *Aphasiology*, 2(3-4), 309–311.
- Kavé, G., Heled, E., Vakil, E., & Agranov, E. (2011). Which verbal fluency measure is most useful in demonstrating executive deficits after traumatic brain injury? *Journal of Clinical and Experimental Neuropsychology*, 33(3), 358–365.
- Keil, K., & Kaszniak, A. W. (2002). Examining executive function in individuals with brain injury: A review. *Aphasiology*, 16(3), 305–335.
- Kerschensteiner, M., Poeck, K., Huber, W., Stachowiak, F.-J., & Weniger, D. (1978). Die Broca-Aphasie. *Journal of Neurology*, 217(4), 223–242.
- Kertesz, A. (1982). *Western Aphasia Battery*. San Antonio and TX: The Psychological Corporation.

- Kim, H., Kim, J., Kim, D. Y., & Heo, J. (2011). Differentiating between aphasic and nonaphasic stroke patients using semantic verbal fluency measures with administration time of 30 seconds. *European neurology*, *65*(2), 113–117.
- King, K. A., Hough, M., Walker, M., Rastatter, M. P., & Holbert, D. (2006a). Mild traumatic brain injury: Effects on naming in word retrieval and discourse. *Brain Injury*, *20*(7), 725–732.
- King, K. A., Hough, M. S., Vos, P., Walker, M. M., & Givens, G. (2006b). Word retrieval following mild TBI: Implications for categorical deficits. *Aphasiology*, *20*(2–4), 233–245.
- Kleissendorf, B., Jaecks, P., & Stenneken, P. (2008). Quantitative und qualitative Aspekte semantischer Wortflüssigkeit bei Menschen im höheren und hohen Lebensalter mit und ohne Verdacht auf leichte kognitive Beeinträchtigungen. *Spektrum Patholinguistik/ 2*, (p. 171).
- Kononenko, I. (1994). Estimating attributes: Analysis and extensions of relief. In *Proceedings of the European Conference on Machine Learning*.
- Koren, R., Kofman, O., & Berger, A. (2005). Analysis of word clustering in verbal fluency of school-aged children. *Archives of Clinical Neuropsychology*, *20*, 1087–1104.
- Kraus, J. A., Schütze, C., Brokate, B., Kröger, B., & Schwendemann, G. (2005). Discriminant analysis of the cognitive performance profile of MS patients differentiates their clinical course. *Journal of Neurology*, *252*(7), 808–813.
- Kujala, P., Portin, R., & Ruutiainen, J. (1996). Language functions in incipient cognitive decline in multiple sclerosis. *Journal of the neurological sciences*, *141*, 79–86.
- Kuljic-Obradovic, D. C. (2003). Subcortical aphasia: Three different language disorder syndromes? *European Journal of Neurology*, *10*(4), 445–448.
- Laakso, K., Brunnegård, K., Hartelius, L., & Ahlsén, E. (2000). Assessing high-level language in individuals with multiple sclerosis: A pilot study. *Clinical linguistics & phonetics*, *14*(5), 329–349.
- Lacour, A., Seze, J. d., Revenco, E., Lebrun, C., Masmoudi, K., Vidry, E., Rumbach, L., Chatel, M., Verier, A., & Vermersch, P. (2004). Acute aphasia in multiple sclerosis: A multicenter study of 22 patients. *Neurology*, *62*(6), 974–977.
- Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine Learning*, *95*, 161–205.

- Lê, K., Mozeiko, J., & Coelho, C. (2011). Discourse analyses: Characterizing cognitive-communication disorders following TBI. *The ASHA Leader*, (pp. 18–21).
- Lehman Blake, M. (2005). Tangential, egocentric, verbose language: Is it right hemisphere brain damage or normal aging? 35th Clinical Aphasiology Conference.
- Lehman Blake, M., & Tompkins, C. A. (2009). Cognitive-communication disorders resulting from right hemisphere brain damage.
URL <http://www.asha.org/uploadedFiles/public/TESCognitive-CommunicationDisordersfromRightHemisphereBrainDamage.pdf>
- Lehmann, E. L., & Romano, J. P. (2005). *Testing statistical hypotheses*. Springer Texts in Statistics. New York: Springer, third ed.
- Lehr, M., Prud'hommeaux, E. T., Shafran, I., & Roark, B. (2012). Fully automated neuropsychological assessment for detecting mild cognitive impairment. In *Proceedings of Interspeech*.
- Lethlean, J. B., & Murdoch, B. (1997). Performance of subjects with multiple sclerosis on tests of high-level language. *Aphasiology*, 11(1), 39–57.
- Lethlean, J. B., & Murdoch, B. E. (1994). Naming errors in multiple sclerosis: Support for a combined semantic/perceptual deficit. *Journal of Neurolinguistics*, 8(3), 207–223.
- Levelt, W. J. M. (1998). *Speaking: From Intention to Articulation*. Cambridge and Massachusetts and London: MIT Press, 5 ed.
- Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment*. Oxford University Press, USA.
- Lindfors, J. W. (1987). *Children's language and learning*. Prentice-Hall Englewood Cliffs, NJ.
- Maier, A., Haderlein, T., Eysholdt, U., Rosanowski, F., Batliner, A., Schuster, M., & Nöth, E. (2009). PEAKS—a system for the automatic evaluation of voice and speech disorders. *Speech Communication*, 51(5), 425–437.
- Maier, A., Hönig, F., Steidl, S., Nöth, E., Horndasch, S., Sauerhöfer, E., Kratz, O., & Moll, G. (2011). An automatic version of a reading disorder test. *ACM Transactions on Speech and Language Processing (TSLP)*, 7(4), 17.
- Malchus, K., Thiele, K., Jaecks, P., & Stenneken, P. (2012). Emotionales Wohlbefinden als Kontextfaktor in ICF-orientierter Sprachtherapie. BKL-Workshop Klinische Linguistik, 11.-12.05.2012, Berlin.

- Marklund, P., Sikstrom, S., Baath, R., & Nilsson, L.-G. (2009). Age effects on semantic coherence: Latent semantic analysis applied to letter fluency data. In *Proceedings of the Third International Conference on Advances in Semantic Processing*.
- Marsland, S. (2009). *Machine learning: An algorithmic perspective*. Chapman & Hall/CRC.
- Martin, I., & McDonald, S. (2003). Weak coherence, no theory of mind, or executive dysfunction? Solving the problem of pragmatic language disorders. *Brain and Language*, (3), 451–466.
- Max Planck Institute for Psycholinguistics (2011). The Language Archive. Nijmegen. The Netherlands. ELAN Linguistic Annotator, Version 4.1.2, <http://tla.mpi.nl/tools/tla-tools/elan/>.
- Mayr, U. (2002). On the dissociation between clustering and switching in verbal fluency: Comment on Troyer, Moscovitch, Winocur, Alexander and Stuss. *Neuropsychologia*, 40, 562–566.
- McDonald, S. (1993). Viewing the brain sideways? Frontal versus right hemisphere explanations of non-aphasic language disorders. *Aphasiology*, 7(6), 535–549.
- McDonald, S. (2000). Exploring the cognitive basis of right-hemisphere pragmatic language disorders. *Brain and Language*, 75(1), 82–107.
- McDonald, S., & Pearce, S. (1998). Requests that overcome listener reluctance: Impairment associated with executive dysfunction in brain injury. *Brain and language*, 61(1), 88–104.
- Messinis, L., Kosmidis, M. H., Vlahou, C., Malegiannaki, A. C., Gatzounis, G., Dimisianos, N., Karra, A., Kiosseoglou, G., Gourzis, P., & Papatathanasopoulos, P. (2013). Phonological fluency strategy of switching differentiates relapsing-remitting and secondary progressive multiple sclerosis patients. *ISRN neurology*, 10, 1–7.
- Miller, D. H., & Leary, S. M. (2007). Primary-progressive multiple sclerosis. *Lancet Neurology*, 6(10), 903–912.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill series in computer science. New York and London: McGraw-Hill.
- Miyake, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76(2), 165–178.

- Moses, M. S., Sheard, C., & Nickels, L. A. (2007). Insights into recurrent perseverative errors in aphasia: A case series approach. *Aphasiology*, *21*(10-11), 975–1001.
- Müller, S. V., & Münte, T. F. (2000). Störungen von Exekutivfunktionen. In W. Sturm, M. Herrmann, & C. W. Wallesch (Eds.) *Lehrbuch der klinischen Neuropsychologie: Grundlagen, Methoden, Diagnostik, Therapie*, (pp. 480–499). Hogrefe & Huber.
- Myers, P. S. (2001). Toward a definition of RHD syndrome. *Aphasiology*, *15*(10-11), 913–918.
- Nadeau, S. E., & Crosson, B. (1997). Subcortical aphasia. *Brain and Language*, *58*(3), 355–402.
- Nickels, L., & Howard, D. (1995). Aphasic naming: What matters? *Neuropsychologia*, *33*(10), 1281–1303.
- Nickels, L., & Howard, D. (2004). Dissociating effects of number of phonemes, number of syllables, and syllabic complexity on word production in aphasia: It's the number of phonemes that counts. *Cognitive Neuropsychology*, *21*(1), 57–78.
- Nilsson, N. J. (1998). Introduction to machine learning: An early draft of a proposed textbook.
URL <http://robotics.stanford.edu/~nilsson/MLBOOK.pdf>
- Noventini, U., Rossini, P. M., Carlesimo, G. A., Graceffa, A., Grasso, M. G., Lupoi, D., Oliveri, M., Orlacchio, A., Pozzilli, C., Rizzato, B., & Caltagirone, C. (2001). Patterns of cognitive impairment in secondary progressive stable phase of multiple sclerosis: Correlations with MRI findings. *European Neurology*, *45*(1), 11–18.
- Norman, D. A., & Shallice, T. (1980). Attention to action: Willed and automatic control of behavior. Tech. rep., DTIC Document.
- Nutter-Upham, K. E., Saykin, A. J., Rabin, L. A., Roth, R. M., Wishart, H. A., Pare, N., & Flashman, L. A. (2008). Verbal fluency performance in amnesic MCI and older adults with cognitive complaints. *Archives of Clinical Neuropsychology*, *23*(3), 229–241.
- Östberg, P., Crinelli, R. M., Danielsson, R., Wahlund, L.-O., Bogdanovic, N., & Fernaeus, S.-E. (2007). A temporal lobe factor in verb fluency. *Cortex*, *43*, 607–615.
- Östberg, P., Fernaeus, S.-E., Hellström, A., Bogdanovic, N., & Wahlund, L.-O. (2005). Impaired verb fluency: A sign of mild cognitive impairment. *Brain and Language*, *95*, 273–279.

- Parrish, C., Roth, C., Roberts, B., & Davie, G. (2009). Assessment of cognitive-communicative disorders of mild traumatic brain injury sustained in combat. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders*, 19(2), 47–57.
- PASW (2009). PASW Statistics 18; statistiksoftware (SPSS).
- Paulsen, J. S., Romero, R., Chan, A., Davis, A. V., Heaton, R. K., & Jeste, D. V. (1996). Impairment of the semantic network in schizophrenia. *Psychiatry research*, 63(2-3), 109–121.
- Pekkala, S. (2012). Verbal fluency tasks and the neuropsychology of language. *The Handbook of the Neuropsychology of Language, Volume 1&2*, (pp. 619–634).
- Pekkala, S., Albert, M., Spiro III, A., & Erkinjuntti, T. (2008). Perseveration in Alzheimer's Disease. *Dementia and geriatric cognitive disorders*, 25(2), 109–114.
- Perani, D., Cappa, S. F., Tettamanti, M., Rosa, M., Scifo, P., Miozzo, A., Basso, A., & Fazio, F. (2003). A fMRI study of word retrieval in aphasia. *Brain and Language*, 85(3), 357–368.
- Perkins, M. R. (2005). Pragmatic ability and disability as emergent phenomena. *Clinical linguistics & phonetics*, 19(5), 367–377.
- Piatt, A. L., Fields, J. A., Paolo, A. M., Koller, W. C., & Tröster, A. I. (1999a). Lexical, semantic, and action verbal fluency in Parkinson's Disease with and without dementia. *Journal of Clinical and Experimental Neuropsychology*, 21(4), 435–443.
- Piatt, A. L., Fields, J. A., Paolo, A. M., & Tröster, A. I. (1999b). Action (verb naming) fluency as an executive function measure: Convergent and divergent evidence of validity. *Neuropsychologia*, 37(13), 1499–1503.
- Piatt, A. L., Fields, J. A., Paolo, A. M., & Tröster, A. I. (2004). Action verbal fluency normative data for the elderly. *Brain and Language*, 89(3), 580–583.
- Pinel, J. P., & Pauli, P. (2007). *Biopsychologie*. München: Pearson Studium.
- Platt, J. (1998). Fast training of Support Vector Machines using sequential minimal optimization. In B. Schoelkopf, C. Burges, & A. Smola (Eds.) *Advances in Kernel Methods - Support Vector Learning*. MIT Press.
- Poeck, K., & Hacke, W. (2001). *Neurologie*. Springer.
- Poeck, K., Kerschensteiner, M., Stachowiak, F. J., & Huber, W. (1974). Die Amnestische Aphasie. *Journal of Neurology*, 207(1), 1–17.

- Possin, K. L., Vincent Filoteo, J., Roesch, S. C., Zizak, V., Rilling, L. M., & Davis, J. D. (2005). Is a perseveration a perseveration? An evaluation of cognitive error types in patients with subcortical pathology. *Journal of Clinical and Experimental Neuropsychology*, 27(8), 953–966.
- Prigatano, G. P. (1986). *Neuropsychological rehabilitation after brain injury*. Baltimore: Johns Hopkins University Press.
- Prud'hommeaux, E., Mitchell, M., & Roark, B. (2011). Using patterns of narrative recall for improved detection of mild cognitive impairment. *Age*, 81, 79–7.
- Prud'hommeaux, E. T., & Roark, B. (2011). Alignment of spoken narratives for automated neuropsychological assessment. In *Automatic Speech Recognition and Understanding (ASRU), 2011 IEEE Workshop on*, (pp. 484–489). IEEE.
- Psyrembel (2002). *Klinisches Wörterbuch*. de Gruyter. Berlin, New York, 259 ed.
- Pusswald, G., & Vass, K. (2011). Multiple Sklerose. In J. Lehrner, G. Pusswald, E. Fertl, W. Strubreither, & I. Kryspin-Exner (Eds.) *Klinische Neuropsychologie*, (pp. 287–302). Wien: Springer Verlag.
- Quintana, M., Guàrdia, J., Sánchez-Benavides, G., Aguilar, M., Molinuevo, J. L., Robles, A., Barquero, M. S., Antúnez, C., Martínez-Parra, C., Frank-García, A., Fernández, M., Blesa, R., & Peña-Casanova, J. (2012). Using artificial neural networks in clinical neuropsychology: High performance in mild cognitive impairment and alzheimer's disease. *Journal of clinical and experimental neuropsychology*, 34(2), 195–208.
- Raboutet, C., Sauzéon, H., Corsini, M.-M., Rodrigues, J., Langevin, S., & N'Kaoua, B. (2010). Performance on a semantic verbal fluency task across time: Dissociation between clustering, switching, and categorical exploitation processes. *Journal of Clinical and Experimental Neuropsychology*, 32(3), 268–280.
- Rao, S. M. (1986). Neuropsychology of multiple sclerosis: A critical review. *Journal of Clinical and Experimental Neuropsychology*, 8(5), 503–542.
- Rao, S. M., Cognitive Function Study Group, & National Multiple Sclerosis Society (1990). *Manual for the Brief Repeatable Battery of Neuropsychological Tests in Multiple Sclerosis*. New York and NY and USA: National Multiple Sclerosis Society.
- Rao, S. M., Grafman, J., DiGiulio, D., Mittenberg, W., Bernardin, L., Leo, G. J., Luchetta, T., & Unverzagt, F. (1991a). Memory dysfunction in

LITERATURVERZEICHNIS

- multiple sclerosis: Its relation to working memory, semantic encoding and implicit learning. *Neuropsychology*, 7(3), 364–374.
- Rao, S. M., Leo, G. J., Ellington, L., Nauertz, T., Bernardin, L., & Unverzagt, F. (1991b). Cognitive dysfunction in multiple sclerosis. II. Impact on employment and social functioning. *Neurology*, 41(5), 692–696.
- Rao, S. M., St. Aubin-Faubert, P., & Leo, G. J. (1989). Information processing speed in patients with multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, 11(4), 471–477.
- Raskin, S. A., & Rearick, E. (1996). Verbal fluency in individuals with mild traumatic brain injury. *Neuropsychology*, 10(3), 416–422.
- Reitan, R. M. (1988). Integration of neuropsychological theory, assessment, and application. *Clinical Neuropsychologist*, 2(4), 331–349.
- Repovš, G., & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience*, 139(1), 5–22.
- Richter, K., Wittler, M., & Hielscher-Fastabend, M. (2006). BIAS-Bielefelder Aphasia Screening. *Zur Diagnostik akuter Aphasien*. Hofheim: NAT-Verlag.
- Roark, B., Mitchell, M., & Hollingshead, K. (2007). Syntactic complexity measures for detecting mild cognitive impairment. In *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*, (pp. 1–8). Association for Computational Linguistics.
- Roark, B., Mitchell, M., Hosom, J., Hollingshead, K., & Kaye, J. (2011). Spoken language derived measures for detecting mild cognitive impairment. *Audio, Speech, and Language Processing, IEEE Transactions on*, 19(7), 2081–2090.
- Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, 126(3), 211–227.
- Rosen, V. M., Sunderland, T., Levy, J., Harwell, A., McGee, L., Hammond, C., Bhupali, D., Putnam, K., Bergeson, J., & Lefkowitz, C. (2005). Apolipoprotein E and category fluency: Evidence for reduced semantic access in healthy normal controls at risk for developing Alzheimer’s disease. *Neuropsychologia*, 43, 647–658.
- Rossell, S. L. (2006). Category fluency performance in patients with schizophrenia and bipolar disorder: The influence of affective categories. *Schizophrenia Research*, 82(2-3), 135–138.

- Rosser, A., & Hodges, J. R. (1994). Initial letter and semantic category fluency in Alzheimer's disease, Huntington's disease, and progressive supranuclear palsy. *Journal of Neurology, Neurosurgery & Psychiatry*, *57*(11), 1389–1394.
- Ruet, A., Deloire, M., Charré-Morin, J., Hamel, D., & Brochet, B. (2013). Cognitive impairment differs between primary progressive and relapsing-remitting MS. *Neurology*.
- Ruff, R. M., Light, R. H., Parker, S. B., & Levin, H. S. (1997). The psychological construct of word fluency. *Brain and Language*, *57*, 394–405.
- Rummer, R. (1996). *Kognitive Beanspruchung beim Sprechen*. Psychologie-Verlag-Union.
- Russell, K. C., Arenth, P. M., Scanlon, J. M., Kessler, L., & Ricker, J. H. (2012). Hemispheric and executive influences on low-level language processing after traumatic brain injury. *Brain injury : [BI]*, *26*(7-8), 984–995.
- Sandson, J., & Albert, M. L. (1984). Varieties of perseveration. *A special issue of Neuropsychologia dedicated to the memory of Henry Hecaen*, *22*(6), 715–732.
- Schade, U. (1992). *Konnektionismus: Zur Modellierung der Sprachproduktion*. Psycholinguistische Studien. Westdeutscher Verlag.
- Schmid, M., Strand, M., Ardal, G., Lund, A., & Hammar, A. (2011). Prolonged impairment in inhibition and semantic fluency in a follow-up study of recurrent major depression. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*, *26*(7), 677–686.
- Schwartz, S., Baldo, J., Graves, R. E., & Brugger, P. (2003). Pervasive influence of semantics in letter and category fluency: A multidimensional approach. *Brain and Language*, *87*(3), 400–411.
- Sepulcre, J., Peraita, H., Goñi, J., Arrondo, G., Martincorena, I., Duque, B., Vélez Mendizábal, N. d., Masdeu, J. C., & Villoslada, P. (2011). Lexical access changes in patients with multiple sclerosis: A two-year follow-up study. *Journal of Clinical and Experimental Neuropsychology*, *33*(2), 169–175.
- Shallice, T. (2002). Fractioning the supervisory system. In D. Stuss, & R. Knight (Eds.) *Principles of frontal lobe functioning*, (pp. 261–277). New York: Oxford University Press.

LITERATURVERZEICHNIS

- Shallice, T., Burgess, P., & Robertson, I. (1996). The domain of supervisory processes and temporal organization of behaviour [and discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 351, 1405–1412.
- Sloetjes, H., & Wittenburg, P. (2008). Annotation by category: Elan and iso dcr. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Sciences*, 283, 1657–1661.
- Sohlberg, M. M., & Mateer, C. A. (2001). *Cognitive rehabilitation: An integrative neuropsychological approach*. New York: Guilford Press.
- Stachowiak, F.-J., Huber, W., Kerschensteiner, M., Poeck, K., & Weniger, D. (1977). Die Globale Aphasie. *Journal of Neurology*, 214(2), 75–87.
- Stuss, D. T. (2011). Traumatic brain injury: Relation to executive dysfunction and the frontal lobes. *Current Opinion in Neurology*, 24(6), 584–589.
- Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological research*, 63(3-4), 289–298.
- Stuss, D. T., & Alexander, M. P. (2007). Is there a dysexecutive syndrome? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481), 901–915.
- Tesak, J. (2006). *Einführung in die Aphasieologie*. Georg Thieme Verlag.
- Testa, J. A., Tröster, A. I., Fields, J. A., Gleason, A. C., Salmon, D. P., & Beatty, W. W. (1998). Semantic fluency performance of patients with cortical and subcortical neurodegenerative diseases. *Aging, Neuropsychology and Cognition*, 5(3), 203–214.
- Thiele, K., Gaspers, J., Cimiano, P., Foltz, A., & Stenneken, P. (under revision). Towards the design of an evaluation method to diagnose cognitive communication disorder: Applying machine learning techniques to word generation data. *Artificial Intelligence in Medicine*.
- Thiele, K., Jaecks, P., Böwering, T., & Stenneken, P. (in Vorbereitung). Analyse von Spontansprache- und Wortgenerierungscharakteristika bei Multipler Sklerose.
- Thiele, K., Malchus, K., Jaecks, P., & Stenneken, P. (2012). Let's talk about emotion! Kommunikation von Emotion mit sprachbeeinträchtigten Patienten. 12. Jahrestagung der GAB and Leipzig.

- Thier, P. (2006). Die funktionelle Architektur des präfrontalen Kortex. In H.-O. Karnath, & P. Thier (Eds.) *Neuropsychologie*, Springer-Lehrbuch, (pp. 471–478). Berlin/Heidelberg: Springer-Verlag.
- Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology*, *14*(2), 167–177.
- Tröster, A. I., Warmflash, V., Osorio, I., Paolo, A. M., Alexander, L. J., & Barr, W. B. (1995). The roles of semantic networks and search efficiency in verbal fluency performance in intractable temporal lobe epilepsy. *Epilepsy Research*, *21*(1), 19–26.
- Troyer, A. K., & Moscovitch, M. (2006). Cognitive processes of verbal fluency tasks. In A. M. Poreh (Ed.) *The quantified process approach to neuropsychological assessment*, Studies on neuropsychology, neurology, and cognition, (pp. 143–160). New York: Taylor & Francis.
- Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138–146.
- Troyer, A. K., Moscovitch, M., Winocur, G., Alexander, M. P., & Stuss, D. (1998). Clustering and switching on verbal fluency: The effects of focal frontal- and temporal-lobe lesions. *Neuropsychologia*, *36*(6), 499–504.
- Turkstra, L., Ylvisaker, M., Coelho, C., Kennedy, M., Sohlberg, M. M., Avery, J., & Yorkston, K. (2005a). Practice guidelines for standardized assessment for persons with traumatic brain injury. *Journal of Medical Speech-Language Pathology*, *13*(2), 9–38.
- Turkstra, L. S., Coelho, C., & Ylvisaker, M. (2005b). The use of standardized tests for individuals with Cognitive-Communication Disorders. *Seminars in speech and language*, *26*(4), 215–222.
- Turkstra, L. S., McDonald, S., & Kaufmann, P. M. (1995). Assessment of pragmatic communication skills in adolescents after traumatic brain injury. *Brain Injury*, *10*(5), 329–345.
- Unsworth, N., Spillers, G. J., & Brewer, G. A. (2011). Variation in verbal fluency: A latent variable analysis of clustering, switching, and overall performance. *The Quarterly Journal of Experimental Psychology*, *64*(3), 447–466.
- Uomoto, J. M. (1991). Evaluation of neuropsychological status after traumatic brain injury. In D. R. Beukelman, & K. M. Yorkston (Eds.) *Communication disorders following traumatic brain injury: Management of cognitive, language, and motor impairments*, (pp. 75–101). Pro-Ed.

LITERATURVERZEICHNIS

- Vogel, A. P., Maruff, P., & Morgan, A. T. (2010). Evaluation of communication assessment practices during the acute stages post stroke. *Journal of Evaluation in Clinical Practice*, *16*(6), 1183–1188.
- Wallesch, C.-W. (2006). Sprache. In H.-O. Karnath, & P. Thier (Eds.) *Neuropsychologie*, Springer-Lehrbuch, (pp. 521–524). Berlin/Heidelberg: Springer-Verlag.
- Wehmeyer, M., & Grötzbach, H. (2006). *Aphasie: Wege aus dem Sprachdschungel*. Berlin: Springer, 3 ed.
- Weniger, D. (2006). Aphasien. In H.-O. Karnath, & P. Thier (Eds.) *Neuropsychologie*, Springer-Lehrbuch, (pp. 356–372). Berlin/Heidelberg: Springer-Verlag.
- Wiig, E. H., & Secord, W. (1989). *Test of Language Competence-Expanded Edition (TLC-Expanded)*. Harcourt Assessment.
- Wilson, D. (2005). New directions for research on pragmatics and modularity. *Lingua*, *115*(8), 1129–1146.
- Wilson, S. M., Ogar, J. M., Laluz, V., Growdon, M., Jang, J., Glenn, S., Miller, B. L., Weiner, M. W., & Gorno-Tempini, M. L. (2009). Automated MRI-based classification of primary progressive aphasia variants. *NeuroImage*, *47*(4), 1558–1567.
- Wishart, H., & Sharpe, D. (1997). Neuropsychological aspects of multiple sclerosis: A quantitative review. *Journal of Clinical and Experimental Neuropsychology*, *19*(6), 810–824.
- Wong, M. N., Murdoch, B., & Whelan, B.-M. (2010). Language disorders subsequent to mild traumatic brain injury (MTBI): Evidence from four cases. *Aphasiology*, *24*(10), 1155–1169.
- Ylvisaker, M., & Szekeres, S. F. (1989). Metacognitive and executive impairments in head-injured children and adults. *Topics in Language Disorders*, *9*(2), 34–49.
- Zakzanis, K. K. (2000). Distinct neurocognitive profiles in multiple sclerosis subtypes. *Archives of Clinical Neuropsychology*, *15*(2), 115–136.
- Ziegler, W., & Zierdt, A. (2008). Telediagnostic assessment of intelligibility in dysarthria: A pilot investigation of MVP-online. *Journal of Communication Disorders*, *41*(6), 553–577.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig verfasst, keine anderen als die angegebenen Hilfsmittel verwendet und wörtlich oder inhaltlich übernommene Stellen als solche kenntlich gemacht habe.

Kristina Thiele
Mannheim, im Mai 2013