Neurokognitive Dynamik semantischer Gedächtnisprozesse in algorithmischen Modellen

Synopsis zur kumulativen Habilitation

Fakultät für Human- und Sozialwissenschaften

Bergische Universität Wuppertal

Dr. phil. Dipl.-Psych. Markus J. Hofmann

Geboren am 2.5.1975 in Würzburg

12.5.2021

Die Habilitationsschrift kann wie folgt zitiert werden:

urn:nbn:de:hbz:468-20210608-090126-6
[http://nbn-resolving.de/urn/resolver.pl?urn=urn%3Anbn%3Ade%3Ahbz%3A468-20210608-090126-6]

DOI: 10.25926/h0d8-f009
[https://doi.org/10.25926/h0d8-f009]

Inhaltsverzeichnis

Gedruckt auf alterungsbeständigem Papier gemäß DIN ISO 9706

0. Zusammenfassung

Um assoziative Verknüpfungen zwischen Wörtern zu definieren, bedient sich die psychologische Gedächtnisforschung immer häufiger algorithmischer Modelle, die das semantische Langzeitgedächtnis aus exemplarischen Textkorpora als „Erfahrungsgrundlage“ ableiten und damit quantitative Vorhersagen von menschlichen Gedächtnisleistungen ermöglichen. Nach einem kurzen Überblick über die Geschichte der Assoziationsforschung von den präquantitativen zu quantitativen Formulierungen der Assoziationsgesetze in Kapitel 1 werden in dieser kumulativen Habilitation verschiedene quantitative Definitionen der Gesetze der Kontiguität, Frequenz und Ähnlichkeit auf deren Fähigkeit getestet, Verhaltens- und neurokognitive Daten vorherzusagen. Kapitel 2 beginnt mit einem Überblick über *interactive activation models* (IAMs), die aus den Ebenen visueller Eigenschaften, Buchstaben und orthographischer Wortformen bestehen, und die darauf aufbauende semantische Ebene im *associative read-out model* (AROM, Hofmann & Jacobs, 2014). Wir diskutieren verschiedene klassische Ansätze, „semantische” Verknüpfungen zu definieren, die das menschliche Verhalten zirkulär auf Basis anderen menschlichen Verhaltens vorhersagen, wie zum Beispiel freie Assoziationen oder Lückentextergänzungswahrscheinlichkeiten (LTEW). Diese Ansätze vergleichen wir insbesondere mit einer algorithmisch definierten Assoziationsstärke (AS), bei der zwei Wörter als „assoziiert” definiert werden, wenn sie signifikant häufiger gemeinsam in den Sätzen eines großen Korpus auftreten, als die Einzelauftretenshäufigkeiten erwarten lassen würden. Schließlich erörtern wir verschiedene Vorhersagen, die sich für neurokognitive Daten aus dem AROM ableiten lassen: Beispielsweise zeigen wir AS-Effekte zwischen den Nomen von Komposita im linken inferioren Frontalgyrus mittels funktioneller Magnetresonanztomographie (fMRT); wir erklären, warum die Erinnerungsraten an Wörter, die in einer Studierphase vorher gelernt wurden, stärker variieren als die Fehlerinnerungen an nicht-studierte Wörter; wir demonstrieren, dass Fehlerinnerungen häufiger auftreten, wenn die Wörter mehr assoziative Verknüpfungen zu den anderen Reizen in solchen episodischen Gedächtnisaufgaben aufweisen; und wie die Anzahl assoziativer Verknüpfungen die Effekte positiver, aber nicht negativer Valenz von Wörtern erklärt. In Kapitel 3 verwenden wir einen klassischen Ansatz, die LTEW, und zeigen mittels funktioneller Nah-Infrarotspektroskopie Effekte der Wortvorhersehbarkeit im okzipitalen Cortex, die einen *Top-down*-

Informationsfluss von der semantischen bis zur Ebene visueller Eigenschaften nahelegen. Ebenso zeigen wir Erwartungsverletzungssignale bei seltenen Wörtern und eine Tendenz für Erwartungsbestätigung bei häufigen Wörtern im orbitofrontalen Cortex (Hofmann et al., 2014). In Kapitel 4.1 soll die AS assoziative Urteile von Wortpaaren auf einer siebenstufigen *Rating*-Skala vorhersagen (Hofmann et al., 2018). Während die AS und die Interaktion der emotionalen Valenzen der beiden Wörter reproduzierbare Varianz in den geplanten Analysen der drei Studien aufklärt, erklärt diese Interaktion keine Varianz mehr, wenn man ein *Skip-gram*-Modell in explorativen Analysen hinzunimmt. Hieraus lässt sich schließen, dass Informationen über den emotionalen Gehalt von Wörtern in den semantischen Strukturen dieses Modells enthalten sind. AS und *Skip-gram*-Modell erklären in Hofmann et al. (2018) jeweils etwa 40 % und 50 % der *Item-level*-Varianz. Die anderen fünf algorithmischen Ansätze konnten keinen zusätzlichen Varianzanteil reproduzierbar aufklären. In Kapitel 4.2 nutzen wir drei algorithmische Modelle zuerst für die Vorhersage von LTEW, dann vergleichen wir alle vier operationalen Definitionen für Vorhersagen beim Satzlesen: Hier wird zum einen mittels Blickbewegungsmessung die Zeit vorhergesagt, die das erfolgreiche Erkennen eines Wortes auf einen Blick benötigt, und zum anderen ereigniskorrelierte Potentiale (EKPs) ab 300 ms nach Reizpräsentation (Hofmann, Biemann, & Remus, 2017). Für diese Vorhersagen sind *Topic*-Modelle weniger gut geeignet, während rekurrente neuronale Netzwerke (RNNs) und *N-gram*-Modelle ähnlich gute Vorhersagen machen. Wir berichten in Kapitel 4.3, dass die Anzahl assoziierter Wörter in einer episodischen Gedächtnisaufgabe EKP-Effekte der semantischen Integration ab 320 ms auslösen. Darüber hinaus legen die EKP-Effekte ab 150 ms nahe, dass die assoziativen Verknüpfungen zu den anderen Wörtern auch den *top-down* getriebenen Zugriff auf ein hypothetisches orthographisches Lexikon beeinflussen (Stuellein, Radach, Jacobs, & Hofmann, 2016). In Kapitel 4.4 demonstrieren wir an Hand eines Vergleichs von 200 ms und 1000 ms zwischen Bahnungs- und Zielwort, der *stimulus onset asynchrony* (SOA), dass die AS Reaktionszeit-Effekte bei beiden SOAs und die Anzahl gemeinsamer Assoziierter Effekte bei kurzer SOA in Wort-/Nichtwort-Entscheidungsaufgaben vorhersagen können (Roelke, Franke et al., 2018). Dieses Befundmuster zeigt, dass sich assoziative und semantische Priming-Effekte aus der klassischen Literatur durch die AS-basierten Definitionen abbilden lassen. Darüber hinaus bestätigen sich einige in Hofmann und Jacobs (2014) vorgeschlagene Vorhersagen für fMRT-Daten (vgl. Kapitel 2): Wir finden Effekte gemeinsamer Assoziierter in okzipitalen, fusiformen, temporalen und inferior frontalen Arealen (Roelke, Franke, Radach,

Jacobs, & Hofmann, 2016). Kapitel 5 bietet deutschsprachige Einführungen in die neurokognitive Modellierung (Kapitel 5.1; Jacobs & Hofmann, 2013) und die Leseforschung mit IAMs (Kapitel 5.2; Radach & Hofmann, 2016). In letzterem Buchkapitel findet sich auch eine exemplarische AROM-Simulation assoziativer Aktivierungsausbreitung über Wörter im Satzkontext. Der Ausblick in Kapitel 6.1 zeigt Simulationen der Dynamik assoziativer Aktivierungsausbreitung, um damit assoziative und semantische Priming-Effekte sowie den Effekt der SOA aus Kapitel 4.4 zu erklären. Dann wird in Kapitel 6.2 die zukünftige Erweiterung des AROMs um eine episodische Ebene diskutiert, welche das Knüpfen neuer Assoziationen im *dentate gyrus* des Hippocampus und Mustertrennung im Sinne des Assoziationsgesetzes des Kontrastes abbildet (vgl. Hofmann & Kuchinke, 2015). In Kapitel 6.3 betrachten wir die in Kapitel 4.2 präsentierten Simulationen einer Lückentextergänzungsaufgabe als typische Aufgabe eines Intelligenztests und diskutieren *N-gram*-Modelle als kristalline Gedächtnismodelle, während RNN-Modelle auch generalisieren können und damit näher an den Begriff der fluiden Intelligenz rücken. Dann erörtern wir den möglichen Beitrag künstlicher Intelligenzen für die Theoriebildung und Praxis der Intelligenzforschung in der Psychologie, bevor wir in den Schlussfolgerungen in Kapitel 7 alle Ergebnisse im Kontext der vier Assoziationsgesetze zusammenfassen.

Schlagwörter: Artificial intelligence, associative read-out model, big data, complementary learning systems, event-related potentials, functional magnetic resonance imaging, functional near-infrared spectroscopy, optical imaging, interactive activation model, recurrent neural network.

0.1 Verzeichnis der Publikationen

Artikel in expertenbegutachteten Fachzeitschriften:

Hofmann, M. J., & Jacobs, A. M. (2014). Interactive activation and competition models and semantic context: From behavioral to brain data. *Neuroscience and Biobehavioral Reviews*, *46*, 85–104. (Publikation 1; Kapitel 2).

Hofmann, M. J., Dambacher, M., Jacobs, A. M., Kliegl, R., Radach, R., Kuchinke, L., Plichta, M. M., Fallgatter, A. J., Herrmann, M. J. (2014). Occipital and orbitofrontal hemodynamics during naturally paced reading: An fNIRS study. *NeuroImage*, *94*, 193–202. (Publikation 2; Kapitel 3).

Hofmann, M. J., Biemann, C., Westbury, C. F., Murusidze, M., Conrad, M., & Jacobs, A. M. (2018). Simple co-occurrence statistics reproducibly predict association ratings. *Cognitive Science*, *42*, 2287–2312. (Publikation 3; Kapitel 4.1).

Stuellein, N., Radach, R. R., Jacobs, A. M., & **Hofmann**, M. J. (2016). No one way ticket from orthography to semantics in recognition memory: N400 and P200 effects of associations. *Brain Research*, *1639*, 88–98. (Publikation 5; Kapitel 4.3).

Roelke, A., Franke, N., Biemann, C., Radach, R., Jacobs, A. M., & **Hofmann**, M. J. (2018). A novel co-occurrence-based approach to predict pure associative and semantic priming. *Psychonomic Bulletin and Review*, *25*(4), 1488–1493. (Publikation 6; Kapitel 4.4).

Hofmann, M. J., & Kuchinke, L. (2015). "Anything is good that stimulates thought" in the hippocampus: Comment on "The quartet theory of human emotions: An integrative and neurofunctional model" by S. Koelsch et al. *Physics of Life Reviews*, *13*, 58–60. (Publikation 9; Kapitel 6).

Weitere Beiträge in wissenschaftlichen Organen und Sammelwerken:

Hofmann, M. J., Biemann, C., & Remus, S. (2017). Benchmarking n-grams, topic models and recurrent neural networks by cloze completions, EEGs and eye movements. In B. Sharp, F. Sedes, & W. Lubaszewsk (Eds.), *Cognitive Approach to Natural Language*

Processing (pp. 197–215). London, UK: ISTE Press Ltd, Elsevier. (Publikation 4; Kapitel 4.2).

Jacobs, A. M., & **Hofmann**, M. J. (2013). Neurokognitive Modellierung. In E. Schröger & S. Koelsch (Eds.), *Enzyklopädie der Psychologie. Affektive und kognitive Neurowissenschaft* (pp. 431–447). Göttingen: Hogrefe. (Publikation 7; Kapitel 5.1).

Radach, R., & **Hofmann**, M. J. (2016). Graphematische Verarbeitung beim Lesen von Wörtern. In U. Domahs & B. Primus (Eds.), *Laut, Gebärde, Buchstabe (Handbuch Sprachwissen, Band 2)* (pp. 455–473). Berlin: De Gruyter Mouton. (Publikation 8; Kapitel 5.2).

0.2 Abkürzungsverzeichnis

ACC – Anterior cingulate cortex

AI – Artificial intelligence

AMSS – Associative memory signal strength

AROM – Associative read-out model

AS – Assoziationsstärke

ATL – Anteriorer Temporallappen

CA3 – Subfeld 3 im cornu ammonis

CBOW – Continuous bag-of-words

CLS – Complementary learning systems

DG – Dentate gyrus

DLPFC – Dorsolateraler präfrontaler Cortex

EEG – Elektroenzephalogramm

EKP – Ereigniskorreliertes Potential

fMRT – Funktionelle Magnetresonanztomographie

FFG – Linker fusiformer Gyrus

fNIRS – Funktionelle Nah-Infrarotspektroskopie

frNIRS – Fixationsrelatierte Nah-Infrarotspektroskopie

GLA – Global lexical activation

IAM – Interactive activation model

IFG – Linker inferiorer Frontalgyrus

IST 2000R – Intelligenz-Struktur-Test 2000R

LDA - Latent dirichlet allocation

LSA – Latent semantic analysis

LTEW – Lückentextergänzungswahrscheinlichkeiten

LZG – Langzeitgedächtnis

MROM – Multiple read-out model

OC – Okzipitaler Cortex

OFC – Orbitofrontaler Cortex

PMI – Pointwise mutual information

PSC – Potsdam Satzcorpus

RNN – Rekurrentes neuronales Netzwerk

RSVP – Rapid serial visual presentation

SFD – Single-fixation duration

SOA – Stimulus onset asynchrony

STL – Superiorer Temporallappen

SRN – Simple recurrent network

XOR – Exclusive-Or

1. Einleitung

Die Bedeutung von Wörtern ist ein grundlegender Bestandteil der meisten psychologischen Theorien. Ebenso basieren die meisten diagnostischen Instrumente der psychologischen Praxis auf Sprache. Somit sind Wörter ein allgegenwärtiges Medium, um damit Bedeutungszusammenhänge zu vermitteln und deren Effekt auf das semantische System im Menschen zu untersuchen. Doch wie genau das semantische System eines Menschen aus seiner Erfahrung konsolidiert wird und wie diese konsolidierten Gedächtnisstrukturen das Verhalten in psychologischen Aufgaben determinieren, das ist meines Erachtens nur unzureichend gut verstanden.

Die Aufgabe der allgemeinen Psychologie besteht darin, grundlegende Theorien und Modelle zu generieren, um damit möglichst allgemeingültige Gesetzmäßigkeiten abzubilden. Da sich immer mehr Psychologen als Naturwissenschaftler begreifen, steht die Psychologie als Wissenschaft vor der Herausforderung, die Errungenschaften im Bereich präquantitativer Theoriebildung in mathematischen und algorithmischen Modellen quantitativ umzusetzen. Wenn wir uns fragen, wie quantitative Theoriebildung im Allgemeinen funktioniert, dann bringt der theoretische Physiker und Nobelpreisträger Richard Feynman (1968[1]) die Antwort hierauf den Punkt (vgl. Popper, 1935):

„*Nun werde ich diskutieren, wie wir nach einem neuen Naturgesetz suchen würden. Im Allgemeinen* ***suchen wir nach einem neuen Naturgesetz*** *durch den folgenden Prozess:* ***Zuerst schätzen wir es****. (...) Dann* ***berechnen wir die Konsequenzen aus dieser Schätzung*** *(...), um zu sehen, was es implizieren würde – und dann* ***vergleichen wir diese berechneten Ergebnisse mit der Natur*** *oder wir sagen wir vergleichen sie mit dem Experiment oder der Erfahrung – vergleichen es direkt mit der Beobachtung, um zu sehen, ob es funktioniert.* ***Wenn es mit dem Experiment nicht übereinstimmt, ist es falsch****! Und diese einfache Aussage ist der Schlüssel zur Naturwissenschaft.“*

[1] „Now I am going to discuss how we would look for a new law. In general, we look for a new law by the following process: First, we guess it. (…) Then we compute the consequences of the guess (…) to see what it would imply – and then we compare those computation results to nature, or we say compare to experiment or experience – compare it directly with the observation to see if it works. If it disagrees with [the] experiment, it’s wrong. And that simple statement is the key to science.”

Diese Habilitation beschäftigt sich mit der Frage, wie sich Assoziationen zwischen Wörtern im semantischen Langzeitgedächtnis (LZG) am besten quantitativ in algorithmischen Modellen abbilden lassen. Um einen tieferen Einblick in die verschiedenen operationalen Definitionen „semantischer" Assoziationen und deren Entstehungsgeschichte zu gewinnen, beginnen wir mit einem kurzen historischen Überblick, der die Entwicklung der psychologischen Forschung zu den „Assoziationsgesetzen" in Kapitel 1.1 aufzeigen soll. Da dieses Forschungsfeld viel zu umfassend ist, um es vollständig an dieser Stelle beschreiben zu können, habe ich einige zentrale Meilensteine ausgewählt, die meine Forschung motiviert haben, und die einen selektiven Überblick über die Entwicklung dieses Forschungsfeldes geben sollen. Im Wesentlichen lassen sich zwei Phasen unterscheiden: Bis zur kognitiven Wende in den 1950er Jahren konnte dieses komplexe Forschungsfeld quasi ausschließlich durch präquantitative Theoriebildung vorangetrieben werden. Ab der kognitiven Wende stand durch die Erfindung des Computers endlich genügend Rechenleistung zur Verfügung um diese Theorien quantitativ abzubilden. Somit konnte getestet werden, was aus den verschiedenen quantitativen Operationalisierungen „semantischer" Relationen folgen würde.

1.1 Eine kurze Geschichte der „Assoziationsgesetze"

Das allgemeine Wissen um „Assoziationsgesetze" geht wahrscheinlich auf Aristoteles Aufsatz „*Über das Gedächtnis und die Erinnerung*" zurück (McKeon, 1941): So schrieb er zum Beispiel „*wir jagen einer Reihe* [an Erinnerungen] *nach, (...) wenn wir in Gedanken (...) bei etwas Ähnlichem oder Gegensätzlichem zu dem beginnen, was wir suchen, oder bei etwas, das mit ihm nahe beisammen ist*" (McKeon, 1941, p. 612[2]). Später beschrieb Hermann Ebbinghaus (1885) – einer der Gründerväter der Psychologie als empirische Wissenschaft – die Gesetze der Kontiguität und der Frequenz so:

„*Von diesen "Gesetzen" nun – wenn man sich mit dem Sprachgebrauch und hoffentlich in Anticipation der Zukunft die Anwendung eines hohen Wortes auf Formeln von ziemlich vagem Charakter gestattet – von diesen Gesetzen ist eines niemals bestritten und angezweifelt worden (...): Vorstellungen, welche gleichzeitig oder in unmittelbarer Aufeinanderfolge in demselben Bewußtsein erzeugt wurden, reproducieren sich gegenseitig, und zwar mit größerer Leichtigkeit in der Richtung der ursprünglichen Folge, und mit um so größerer Sicherheit, je häufiger sie beisammen waren.*" (Ebbinghaus, 1885, p. 124).

Olson und Hergenhahn (2017, p. 30[3]) fassten die vier Assoziationsgesetze so zusammen: Die Erfahrung oder Erinnerung eines Objektes löst die Erinnerung an Dinge aus, …

(i) die mit dem Objekt in der Vergangenheit beisammen waren (Gesetz der Kontiguität);
(ii) insbesondere wenn dies häufig geschah (Gesetz der Frequenz);
(iii) die dem Objekt ähnlich sind (Gesetz der Ähnlichkeit)
(iv) oder gegensätzlich (Gesetz des Kontrastes).

Der nächste zentrale Meilenstein der psychologischen Assoziationsforschung bestand darin, individuelle semantische Strukturen zu untersuchen: Carl Gustav Jung verwendete bereits Anfang des 20. Jahrhunderts die freie Assoziationsaufgabe (Jung-Merker & Rüf, 2011): Es wurde ein Wort präsentiert, und die Probanden sollten das erste Wort nennen, das ihnen dazu

[2] „we hunt up the series (…) having started in thought (…) from something either similar, or contrary, to what we seek or else from the thing contiguous with it".
[3] „ (…) Aristotle formulated his laws of association. He said that the experience or recall of one object will tend to elicit the recall of things similar to that object (law of similarity), recall of opposite things (law of contrast), or recall of things that were originally experienced along with that object (law of contiguity). Aristotle also noted that the more frequently two things are experienced together, the more likely it will be that the experience or recall of one will stimulate the recall of the second. Later in history this came to be known as the law of frequency".

einfällt. Er zeigte zuerst allgemeine Befunde einer psychisch unauffälligen (Norm-)Stichprobe, wie zum Beispiel, dass die „*Konkreta*" kürzere Reaktionszeiten auslösen als „*allgemeine Begriffe*" (Jung-Merker & Rüf, 2011, p. 286; vgl. Westbury et al., 2013). Dann identifizierte er diese Aufgabe als ein Instrument, um psychische Störungen zu diagnostizieren: „*Die über dem wahrscheinlichen Mittel liegenden Reaktionszeiten sind zum größeren Teil verursacht durch das Auftreten von intensiven Gefühl*[s]*störungen, welche individuell wichtigen Vorstellungskomplexen angehören. Der Grund der Zeitverlängerung ist momentan meist nicht bewusst. Die zu langen Reaktionszeiten können daher als Mittel zur Auffindung affektbetonter (auch unbewußter) Vorstellungskomplexe dienen. (Wichtig bei Hysterie!)*" (Jung-Merker & Rüf, 2011, p. 298).

Während die Psychologie außerhalb der vereinigten Staaten, zum Beispiel mit Piaget in Genf oder Luria in Moskau, an solchen „geistigen" Konzepten in der ersten Hälfte des 20. Jahrhunderts weiter festhielt (Miller, 2003), fokussierten sich die Behavioristen in den vereinigten Staaten auf das beobachtbare Verhalten. Inspiriert durch die frühen Arbeiten des Physiologen Pawlow (Thompson, 1902, p. 152[4]), der „*über die unerwartete Art und Weise*" schrieb, „*wie die Psychologie und Physiologie der Speicheldrüsen miteinander assoziiert werden*", betrachtete zum Beispiel Guthrie (1930) die Konditionierung entsprechend dem Gesetz der Kontiguität als Assoziation, die sich zwischen Reiz und Reaktion bei wiederholtem Auftreten ausbildet. Das Gesetz der Frequenz, dass diese Assoziation durch wiederholte Reiz-Reaktions-Kopplung stärker wird, nannte er auch Gesetz der Übung. Im Gegensatz zu diesem Ansatz der klassischen Konditionierung, nahm man bei operanter Konditionierung an, dass die Assoziation zwischen Reiz und Reaktion verstärkt wird, wenn hierauf eine positive Konsequenz folgt oder eine negative Konsequenz wegfällt (z. B. Thorndyke, 1898). In der Endphase des Behaviorismus bot Skinner (1948) in seinem Buch „*Verbal Behavior*" eine Vorhersage, wie sich Jungs Ergebnisse aus der freien Assoziationsaufgabe durch assoziatives Lernen erklären lassen (vgl. Jung-Merker & Rüf, 2011): „*Manche verbale Antworten können nur durch eine kausale Relation zur vorherigen verbalen Stimulation erklärt werden, die entweder vom Sprecher selbst oder von anderen Sprechern stammt. Lasst uns eine Reaktion,*

[4] „Thus, in a quite unexpected way, the physiology and psychology of the salivary glands have come to be associated together".

die durch einen vorherigen sprachlichen Reiz in anderer Form kontrolliert wird, eine intraverbale Reaktion nennen. (...) Wir könnten annehmen (...), dass wenn zwei verbale Formen nahe beieinander im normalen Diskurs auftreten, sie eine intraverbale Verbindung aufbauen" (Skinner, 1948, pp. 46-48[5]). Neben dem hier angesprochenen Gesetz der Kontiguität fand bei Skinner auch das Gesetz der Frequenz Berücksichtigung: „*Die Relationen scheinen die Häufigkeit benachbarter Verwendung abzubilden. Die Wortform „Meer" tritt wahrscheinlich im Kontext von „See" auf, „Tier" im Kontext von „Katze" und „Tränen" im Kontext von „Schmerz"* (Skinner, 1948, p. 48[6]). Die folgende verbale Beschreibung beinhaltete bereits eine probabilistische Formulierung, wie sie zum Beispiel auch die in Kapitel 4.2 präsentierten Modelle der Lückentextergänzungsaufgabe aufweisen werden: „*Verbale Stimuli* (…) *machen bestimmte Formen der Antwort einfach wahrscheinlicher*" (Skinner, 1948, p. 50[7]). Ende der 1940er Jahre wendeten sich auch in Amerika die ersten Forscher gegen das Credo des Behaviorismus, sich ausschließlich auf beobachtbares Verhalten zu fokussieren (Watson, 1913). So versuchte zum Beispiel Donald Hebb (1949), eine Verbindung zwischen menschlichem Verhalten und den physischen Verbindungen zwischen Nervenzellen herzustellen. In seinem 1949 erschienenen Buch „*The organization of behavior*" schlug er das Prinzip vor, das später unter dem Namen „*Hebbian learning*" bekannt wurde (z. B. O'Reilly, 1998): „*Zwei Zellen oder Systeme von Zellen, die wiederholt zur selben Zeit aktiv sind, werden dazu tendieren, „assoziiert" zu werden, so dass die Aktivität in der einen die Aktivität in der anderen Zelle erleichtert*" (Hebb, 1949, p. 70[8], siehe auch p. 62). Hebb (1949, p. 60[9]), schlug vor, dass „*eine wiederholte Stimulation spezifischer Rezeptoren langsam zur Formation einer Gruppe von Assoziations-Bereichs-Zellen führen wird"*, einem *cell assembly*. „*Diese können*

[5] „Some verbal responses can be accounted for only by (...) a causal relation to prior verbal stimulation, arising from the behavior of either the speaker himself or other speakers. Let us call a response which is controlled by a prior verbal stimulus of different form an Intraverbal Response. (...) We may assume (...), that when two verbal forms occur close together in normal discourse they will acquire an intraverbal connection".

[6] „All the other relations appear to approximate the frequencies of contiguous usage. The form sea is likely to occur in the context of lake, animal in the context of cat, and tears in the context of pain".

[7] „Verbal stimuli (…) simply make certain forms of response more likely to occur".

[8] „[A]ny two cells or systems of cells that are repeatedly active at the same time will tend to become "associated," so that activity in one facilitates activity in the other".

[9] „It is proposed first that a repeated stimulation of specific receptors will lead slowly to the formation of an "assembly" of association-area cells which can act briefly as a closed system after stimulation has ceased; this prolongs the time during which the structural changes of learning can occur and constitutes the simplest instance of a representative process (image or idea)".

sich kurz wie ein geschlossenes System verhalten, nachdem die Stimulation beendet ist; dies verlängert die Zeit, während der strukturelle Veränderungen des Lernens auftreten, und konstituiert den einfachsten Fall eines repräsentativen Prozesses (eines Abbildes oder einer Idee)" (Hebb, 1949, p. 60). Unter dem Abschnitt „*Mechanismen des erwachsenen Lernens*" dachte er semantische Konzepte als Interaktion mehrerer solcher Subsysteme und schrieb, dass „*zwei Konzepte, die miteinander assoziiert werden sollen, Phasen gemeinsamer Gruppen-Aktivierungen aufweisen*" (Hebb, 1949, p. 130[10]).

Ein weiterer Denkansatz der 1940er Jahre, der insbesondere aus den Ingenieurswissenschaften hervorgegangen ist, war die Informationstheorie. Gemäß Shannon (1948) war die Information, die von einem Sender zum Empfänger übermittelt wird, mit Rauschen behaftet. Die Informationstheorie bot formal konkrete Werkzeuge an, mit denen sich die Kapazität übermittelter Information in Bit oder die Entropie als Quelle der Unsicherheit der rauschbehafteten Signale zwischen Sender und Empfänger quantifizieren lassen. Von dieser Denkströmung beeinflusste Entwicklungen in der Psychologie waren die Theorie der Signalentdeckung (z. B. Green & Swets, 1966) und der 1955 von George Miller gehaltene Vortrag über die „*magical number seven*" (Miller, 1994): Auf Grund einer Vielzahl verschiedener Gedächtnisexperimente postulierte er, dass die Kapazität des Kurzzeitgedächtnisses etwa sieben sinntragende Einheiten, sogenannte *chunks*, umfasst. Später ging man jedoch davon aus, dass die Zahl sieben sich darauf zurückführen lässt, dass zum Beispiel zwischen Wort-Reizen semantische Überlappungen bestehen, welche das Zusammenfassen zu einer Einheit höherer Ordnung, das sogenannte *chunking* erleichtern: Die tatsächliche Kapazität des Kurzzeitgedächtnisses bei semantisch unabhängigen Einheiten schätzt man heute auf etwa vier unabhängige sinntragende Einheiten (Cowan, 2000).

Später räumte Miller (2003, p.141[11]) ein, dass er in den frühen 1950er Jahren frustriert davon war, Shannons (1948) Theorie der Information auf die Psychologie anzuwenden. Es ließen sich aus der Informationstheorie Assoziationsmaße zwischen Wörtern ableiten, die auch heute noch Verwendung finden (z. B. Westbury et al., 2013; Westbury, Keith, Briesemeister, Hofmann, & Jacobs, 2015). Ein Maß, das die Informationsüberlappung zweier Variablen angibt, war die

[10] „that two concepts to be associated may have phases (assembly actions) in common".
[11] „those years in the early 1950s. (…) During those years I personally became frustrated in my attempts to apply Claude Shannon's theory of information to psychology."

sogenannte *mutual information* (Manning & Schütze, 1999, p. 66f). Wenn man die Einzel- und gemeinsamen Auftretenswahrscheinlichkeiten zweier Wörter vergleicht, spricht man hier von *pointwise mutual information* (PMI), die in Kapitel 4.1 als Assoziationsmaß zwischen zwei Wörtern verwendet wird (Bouma, 2009).

Miller (2003[12]) erinnerte sich an das Jahr 1956 als kritisch für die kognitive Wende und die Idee menschlicher Informationsverarbeitung. In diesem Jahr veranstalteten McCarthy, Minsky, Shannon und Rochester eine Konferenz in Dartmouth über *artificial intelligence* (AI; Minsky, 1961), die von nahezu jedem besucht wurde, der in diesem Feld arbeitete, und zu der fünf Jahre später und nach unzähligen Revisionen Minskys (1961) Aufsatz mit dem Titel „*Steps to artificial intelligence*" erschien. Hier fasste er den State of the Art heuristischer Programmierung zusammen, um damit die innere Struktur kognitiver Systeme zum Beispiel im Hinblick auf Suche, Mustererkennung, Lernen, Planen und induktives Schlussfolgern zu beschreiben (Minsky, 1961).

Schließlich erschien im Jahr 1957 zum einen „*Syntactic structures*" von Noam Chomsky (2002), in dem er die Phrasenstrukturgrammatik als „mentalistische" Theorie der Linguistik präsentierte. Zum anderen erschien kurz darauf seine vielzitierte Rezension von Skinners (1948) „*Verbal behavior*": Chomsky (1959) sah das „*Problem, dass er* [Skinner] *sich selbst auf direkt beobachtbare Variablen limitiert, das heißt Input-Output-Relationen. (…) Man würde natürlich erwarten, dass die Vorhersage des Verhaltens eines komplexen Organismus (oder Maschine) zusätzlich zur Information über die externe Stimulation ebenso Wissen über die interne Struktur des Organismus benötigt*" (p. 27[13]). Chomsky rief dazu auf, einen Blick in die *black box* menschlicher Kognition zu wagen, und für viele leitete er damit die kognitive Wende der amerikanischen Psychologie ein.

---

[12] „In the Historical Addendum to Newell and Simon's Human Problem Solving [3] they say: '1956 could be taken as the critical year for the development of information processing psychology' (p. 878). This is not difficult to justify. 1956 was the year that McCarthy, Minsky, Shannon and Nat Rochester held a conference on artificial intelligence at Dartmouth that was attended by nearly everyone working in the field at that time." (…) Minsky circulated a technical report that, after many revisions, and 5 years later, became his influential article, 'Steps toward artificial intelligence' [5]".

[13] „problem (…) that he limits himself to study of 'observables', i.e. input-output relations. (…) One would naturally expect that prediction of the behavior of a complex organism (or machine) would require, in addition to information about external stimulation, knowledge of the internal structure of the organism".

Mit solchen Ansätzen aus Philosophie, Psychologie, Neurowissenschaften, Computerwissenschaften und Linguistik wurde spätestens 1960 klar, dass „*etwas Interdisziplinäres passierte*" (Miller, 2003, p.143[14]). Jede dieser Wissenschaften sah die Kognition jeweils aus einem anderen Blickwinkel; und jede Wissenschaft kam damit weit genug, um zu erkennen, dass die Lösung einiger ihrer Probleme kritisch davon abhängt, ob die anderen Disziplinen Erfolg haben (Miller, 2003, p.143). So entstand das interdisziplinäre Forschungsfeld der Kognitionswissenschaften. Für die psychologische Theoriebildung erschloss die Erfindung des Computers endlich hinreichend viel Rechen- und Speicherkapazität, um die Assoziationsgesetze operational definieren zu können. Damit konnten die ersten quantitativen Modelle erstellt werden, die assoziative Prozesse abbilden.

Das *perceptron* war ein sehr frühes algorithmisches Modell aus den ersten Jahren nach der kognitiven Wende, welches für die weitere Entwicklung „assoziativer Netzwerke" maßgeblich war (Rosenblatt, 1958). Es verfügte über eine *projection area*, die Reizinformationen von der Retina erhielt und als visuelle Informationen abbildete (vgl. Abb. 1, Rosenblatt, 1959, p. 389). Diese Ebene war mit einer *association area* verbunden. Die darin enthaltenen Einheiten wurden rein zufällig verschaltet und waren mit den verschiedenen *responses* verbunden. Die Einheiten der *association area* wiesen bereits eine Aktivierungsschwelle auf, das heißt, es musste ein spezifischer kritischer Input-Wert in der Summe erreicht werden, bevor eine Einheit in einer *all-or-nothing response* feuerte. Dieses Netzwerk trainierte Rosenblatt (1958) darauf, bestimmte *responses* durch positive und negative Verstärkung zu lernen. Im Prinzip enthielt dieses Modell bereits die wichtigsten Elemente von Modellen, die später als konnektionistische Modelle in der Psychologie (z. B. Grainger & Jacobs, 1998; Page, 2000; Thorndyke, 1898) oder als neuronale Netzwerke im Forschungsfeld des maschinellen Lernens bezeichnet wurden (z. B. LeCun, Bengio, & Hinton, 2015).

[14] „By 1960 it was clear that something interdisciplinary was happening. (…) at least six disciplines were involved: psychology, linguistics, neuroscience, computer science, anthropology and philosophy. I saw psychology, linguistics and computer science as central, the other three as peripheral. (…) Each, by historical accident, had inherited a particular way of looking at cognition and each had progressed far enough to recognize that the solution to some of its problems depended crucially on the solution of problems traditionally allocated to other disciplines".

Im Gegensatz zur *association area* von Rosenblatt (1958), in welcher ein Stimulus verteilt auf subsymbolische Einheiten repräsentiert wurde, begann Ross Quillian (1962, p. 17[15]) damit, „*die Bedeutung der natürlichen Sprache in einem Format zu speichern, das der menschlichen Fähigkeit zu verstehen ähnlich ist*". Auf Basis symbolischer Repräsentationen schrieb er über „*maschinelle Übersetzungsprogramme, die dazu in der Lage sein werden, bestimmte Probleme, wie zum Beispiel die Auflösung von Mehrdeutigkeit, zu lösen*" (Quillian, 1962, p. 17). In seinem 1967 erschienen Artikel „*Word concepts: A theory and simulation of some basic semantic capabilities*" präsentierte er ein „handverdrahtetes" Modell am Beispiel verschiedener Bedeutungen des Begriffes „*plant*", die durch die *type nodes* abgebildet wurden: Die eine *type node* von „*plant*" war mit den *token nodes live* und *animal* verknüpft, die andere mit *apparatus*, *process* und *industry*. Dabei ließen sich die *token nodes* beispielsweise als semantische Eigenschaften betrachten, welche die Bedeutung der *type nodes* definieren. Später konkretisierten Collins und Quillian (1969) die Theorie als hierarchische Form der taxonomischen Wissensrepräsentation, die sie an den Begriffen *„animal"*, *„bird"* und *„canary"* veranschaulichten. Sie schlugen eine ökonomische Repräsentation des Wissens vor, bei der allgemeine Eigenschaften von *„bird"*, wie zum Beispiel *can fly*, auf dieser Ebene gespeichert werden. Dann zeigten sie, dass Versuchspersonen für eine Satz-Verifikationsaufgabe länger benötigten, wenn die Eigenschaften in weiter entfernten Hierarchie-Ebenen gespeichert waren. *A „canary" can fly* ließ sich demnach schneller als wahre Aussage bestätigen als *a „canary" has skin*. Letztere Eigenschaft wäre unter *„animal"* gespeichert und wäre damit mindestens eine Hierarchie-Ebene weiter von *„canary"* entfernt. Schließlich erschien Collins und Loftus' (1975) „*spreading-activation theory of semantic processing*": Hier wurde die assoziative Aktivierungsausbreitung in einem semantischen Netzwerk als dynamischer kognitiver Prozess gesehen, der Phänomene wie Priming erklären sollte (Collins & Loftus, 1975, p. 411[16]): „*Wenn ein Konzept verarbeitet wird, breitet sich die*

[15] „This paper argues that machine translation programs will be able to solve certain problems, e.g., the resolution of polysemy, only by storing the meaning of natural language words in a medium and a format providing properties similar to those of human "understanding"."

[16] „When a concept is processed (or stimulated), activation spreads out along the paths of the network in a decreasing gradient. The decrease is inversely proportional to the accessibility or strength of the links in the path. (…) The longer a concept is continuously processed (either by reading, hearing, or rehearsing it), the longer activation is released from the node of the concept at a fixed rate. Only one concept can be actively processed at a time (…) Activation decreases over time and/or intervening activity. (…) The conceptual (semantic) network is organized along the lines of semantic similarity. The more properties two concepts have in common, the more links there are between the two nodes via these properties and the more closely related are the concepts."

Aktivierung entlang der Pfade des Netzwerks mit einem absteigenden Gradienten aus. Der Abstieg ist invers proportional zur Abrufbarkeit oder zur Stärke der Verbindungen im Pfad. (…) *Je länger das Konzept kontinuierlich verarbeitet wird (entweder durch das Lesen, Hören oder durch Wiederholung), desto länger löst es Aktivierungen von der konzeptuellen Einheit mit einer fixen Rate aus*". Sie gingen davon aus, dass „*zu jeder Zeit nur ein Konzept aktiv sein kann"* und *„die Aktivierung mit der Zeit oder durch intervenierende Aktivität abnimmt*". Ihr konzeptuelles Netzwerk beinhaltete ebenso Vorschläge, wie sich das Gesetz der Ähnlichkeit abbilden lässt: „*Je mehr Eigenschaften zwei Konzepte gemeinsam haben und je mehr Verknüpfungen zwei Konzept-Nodes über diese Eigenschaften aufweisen, desto enger sind diese beiden Konzepte verknüpft*" (Collins & Loftus, 1975, p. 411). Collins und Loftus (1975) boten auf Basis des exemplarischen, handverdrahteten Computermodells von Quillian (1967) eine verbaltheoretische Weiterentwicklung. Doch für eine dynamische Simulation assoziativer Aktivierungsausbreitung im semantischen LZG blieb noch die Frage offen, wie dieses exemplarische, handverdrahtete Modell von Quillian (1967) auf ein „reales semantisches Lexikon" an Weltwissen hochskaliert werden kann.

Während Quillian, Collins und Loftus mit der Perspektive der Verarbeitungsdynamik symbolischer Wort-Einheiten am Beispiel eines Teil-Netzwerks Pionierarbeit leisteten, machten sich Rescorla und Wagner (1972) darüber Gedanken, wie Konditionierung und Reiz-Reaktions-Lernen formal gefasst werden können. Sie beschrieben die Idee einer Asymptote, die man im Rahmen dieser probabilistischen Theorie als eine Art maximaler Lernwahrscheinlichkeit bezeichnen könnte. Wenn man die Veränderung der Erinnerungswahrscheinlichkeit für einen Lern-Durchgang berechnete, musste von dieser Asymptote die Reaktionswahrscheinlichkeit aus dem letzten Durchgang abgezogen werden (Rescorla & Wagner, 1972, p. 75). Dieser Term wurde dann mit der eigentlichen Lernrate multipliziert, welche beispielsweise die verstärkende Wirkung auf die Assoziation durch positive Konsequenzen abbildet. Je höher die Reaktions-Wahrscheinlichkeit auf einen Reiz im letzten Durchgang, desto geringer wurde die Erhöhung der Erinnerungswahrscheinlichkeit im aktuellen Durchgang bei konstanter Lernrate: Je weniger man im letzten Lerndurchgang „wusste", desto mehr kann man aus dem aktuellen Lerndurchgang lernen.

Vermutlich auch weil die Rechen- und Speicherleistungen der Computer in den 1970er und 1980er Jahren für eine realitätsnahe, dynamische Simulation „intraverbaler" Reiz-Reaktions-

Schemata für die Psychologen noch nicht erschwinglich waren (Collins & Loftus, 1975; Skinner, 1948), machte man sich daran, kleinere repräsentative Einheiten zu simulieren, da es davon weniger gab (vgl. das *granularity problem* bei Ziegler & Goswami, 2005). So ließ sich die Aktivierungsausbreitung entlang visueller Eigenschaften, Buchstaben und Wortformen in der von Collins und Loftus (1975) vorgeschlagenen Dynamik simulieren (siehe Abb. 1 in Kapitel 2, McClelland & Rumelhart, 1981). Im *interactive activation model* (IAM) wurden verschiedene formale Annahmen, wie zum Beispiel die asymptotische maximale Aktivierung und die bereits im *perceptron* enthaltene Aktivierungsschwelle (McClelland & Rumelhart, 1981; Rescorla & Wagner, 1972; Rosenblatt, 1958), zusammengeführt und um weitere Annahmen, wie die (Gedächtnis-)Zerfallsrate (*decay*), erweitert (Collins & Loftus, 1975). Ein weiterer Grund, warum die assoziative Aktivierungsausbreitung zwischen semantischen Repräsentationen noch nicht simuliert werden konnte, war sicherlich, dass die Frage nach der assoziativen Verdrahtung des semantischen Gedächtnisses noch immer keine allgemeingültigen, operational definierbaren Antworten für vollständig symbolisch repräsentierende Modelle bereithielt.

Während subsymbolische Modelle, wie das *triangle model,* damit begannen, Lesen und Worterkennung im Spannungsfeld zwischen Orthographie, Phonologie und Semantik zu betrachten (Harm & Seidenberg, 2004; Seidenberg & McClelland, 1989), entstanden auch die ersten Modelle, welche die visuelle Worterkennung beim Benennen von Wörtern auf Basis symbolischer Repräsentationen abbilden. Coltheart, Rastle, Perry, Langdon und Ziegler (2001) ergänzten die direkte Route des lexikalischen Zugriffs, operationalisiert durch das IAM (McClelland & Rumelhart, 1981), später um eine Graphem-Phonem Route (vgl. Coltheart, Curtis, Atkins & Haller, 1993). In den resultierenden *Dual-route*-Modellen blieb das semantische System jedoch eine theoretische Annahme, die noch nicht in konkrete Simulationen umgesetzt werden konnte (vgl. Perry, Ziegler & Zorzi, 2007). Schließlich ergänzten Grainger und Jacobs (1996) das IAM im *multiple read-out model* (MROM) um multiple Entscheidungsmechanismen, die wir in den Simulationen in Kapitel 6.1 genauer kennenlernen werden. Damit machten sie die Simulation der visuellen Worterkennung in Rubensteins lexikalischer Entscheidungsaufgabe möglich (Rubenstein, Garfield, & Millikan, 1970) – derselbe Rubenstein, der kurz vorher bereits vorgeschlagen hatte, wie sich das Gesetz der Ähnlichkeit operational definieren ließe: Semantisch ähnliche Wörter, wie Synonyme,

treten in ähnlichen sprachlichen Kontexten auf (Rubenstein & Goodenough, 1965; vgl., Harris, 1963, p. 15f).

Die in der Psychologie wohl immer noch meistbekannte Methode, solche Kontexte zu definieren, ist die *latent semantic analysis* (LSA, Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Landauer & Dumais, 1997; vgl. Günther, Dudschig, & Kaup, 2015). Als „Kontext" definierten diese Autoren Dokumente. Sie errechneten latente semantische Dimensionen, die das kookkurrente, das heißt gemeinsame Auftreten von Wörtern in Dokumenten bestimmen. Dabei reduziert man in der LSA die Anzahl möglicher semantischer Dimensionen mittels *single-value decomposition* – eine Methode, die später durch die probabilistisch formulierte *latent dirichlet allocation* ersetzt wurde (LDA; Blei, Ng, & Jordan, 2003). Für psychologische Aufgaben konnten Griffiths, Steyvers und Tennenbaum (2007) zeigen, dass viele Ergebnisse besser durch die LDA-basierten *Topic*-Modelle erklärt werden können. In Kapitel 4.2 werden wir ein solches *Topic*-Modell verwenden, um zu testen, ob die aus Dokumenten gewonnene semantische Struktur einen Einfluss auf das Satzlesen hat (Hofmann, Biemann, & Remus, 2017). Ähnlich wie konnektionistische Modelle, welche die Bedeutung eines Objektes verteilt auf subsymbolische Einheiten formal fassten (z. B. Rosenblatt, 1958), definierten die LSA und das *Topic*-Modell die Bedeutung eines Wortes verteilt über latente semantische Variablen. Dies bietet den Vorteil, dass die Bedeutung eines Wortes durch relativ wenige latente Variablen abgebildet werden kann, und damit ein relativ geringer Speicherbedarf für die Repräsentation der Bedeutung eines Wortes entsteht.

Während Deerwester et al. (1990) und Landauer und Dumais (1997) die Bedeutung eines Wortes durch ihre faktorenanalytisch inspirierten Ansätze fassten, wurde innerhalb der Literatur konnektionistischer Theoriebildung mit verteilten Repräsentationen ein Ansatz gefunden, dessen Potential erst in jüngster Zeit mit der Verfügbarkeit leistungsfähiger Computer realisiert wurde (Mikolov, Chen, Corrado, & Dean, 2013). Jeffrey Elman (1990) griff Jordans (1986) Idee der rekurrenten Verschaltung zwischen der verteilten, subsymbolischen Repräsentation des aktuellen Wortes und des vorangegangenen Sprachkontextes auf, um ein Netzwerkmodell darauf zu trainieren, das jeweils nächste Wort im Satzkontext vorherzusagen (vgl. Kapitel 4.1, Exkurs: Rekurrente neuronale Netzwerkmodelle). Er konnte zum Beispiel zeigen, dass sich mit einem solchen relativ einfachen Lernmechanismus sowohl syntaktische

als auch semantische Ähnlichkeiten abbilden lassen, was Chomskys (2002, p. 17[17]) Ansicht in Frage stellt, dass „*Grammatik autonom und unabhängig von der Bedeutung ist*". Dies verdeutlichte Chomsky an dem syntaktisch korrekten, aber semantisch inkorrekten Satz „*Farblose, grüne Ideen schlafen wütend*" (Chomsky, 2002, p. 15[18]). Später wurde die grundlegende Architektur dieser rekurrenten neuronalen Netzwerkmodelle von Mikolov (2012) übernommen und an realistisch großen Textkorpora trainiert. Ähnliche Modellarchitekturen sind heute nicht nur State of the Art, was komputational konkrete Theoriebildung in der Psychologie angeht (z. B. Bhatia, 2017; Hofmann et al., 2017; Mandera, Keulers, & Brysbaert, 2017), sondern auf Basis dieser wurde beispielsweise auch der sprachtechnologische Bereich maschineller Übersetzung deutlich verbessert (z. B. Mikolov, Le & Sutskever, 2013).

Und was wurde später aus Ross Quillians (1962, p. 17) Idee, „*die Bedeutung der natürlichen Sprache in einem Format zu speichern, das der menschlichen Fähigkeit zu verstehen ähnlich ist*"? Der meines Wissens erste Ansatz, eine symbolisch-semantische Ebene in ein konnektionistisches Netzwerk zu implementieren, stammte von Wettler und Rapp (1989). Sie nutzten die Einzel- und gemeinsamen Auftretenswahrscheinlichkeiten von bis zu 270 Wörtern aus der PsycInfo-Datenbank, um die Exzitation in einem assoziativen Netzwerk zu skalieren. Dann zeigten sie, wie eine dynamische assoziative Aktivierungsausbreitung dazu führt, dass auch ähnliche Begriffe aktiv werden können und dass wenig distinkte Begriffe der psychologischen Literatur, wie „Effekt", nach einigen Simulationszyklen eine geringere Aktivierung aufweisen als inhaltlich trennschärfere Wörter. Sie zeigten somit einen Mechanismus auf, der die Suche in großen Datenbanken mittels dynamischer Aktivierungsausbreitung verbessern soll.

Als alternatives Assoziationsmaß zum rein probabilistischen Ansatz von Wettler und Rapp (1989) hat Ted Dunning (1993) einen sehr einfachen Lernmechanismus auf Basis eines *Log-likelihood*-Tests vorgeschlagen. Zwei Wörter wurden als „assoziiert" definiert, wenn sie signifikant häufiger gemeinsam in einem Kontext auftreten, als sich dies durch ihre Einzelauftretenshäufigkeiten per Zufall erwarten ließe. In seinem *Log-likelihood*-Test fanden sich die Gesetze der Kontiguität und Frequenz wieder. Diese wurden dabei aber vollständig symbolisch abgebildet: Ob zwei Wörter „assoziiert" definiert wurden oder nicht, entschied

[17] „grammar is autonomous and independent of meaning".
[18] „Colorless green ideas sleep furiously".

demnach die Signifikanzschwelle. Darüber hinaus fand sich in der resultierenden *log likelihood ratio*, dem χ^2-Wert, auch eine quantitative Skalierung des Gesetzes der Frequenz (vgl. Hofmann et al., 2018).

Als wir das IAM um eine semantische Ebene im *associative read-out model* (AROM), erweiterten (Hofmann, Kuchinke, Biemann, Tamm & Jacobs, 2011), wurde die Interaktion dieser beiden Gesetze in einem einzigen Wert gefasst (Hofmann et al., 2018): Gemäß der Idee frequenzgewichteter Kontiguität treten zwei Wörter signifikant häufig miteinander in den Sätzen eines großen Korpus aus zum Beispiel 43 Millionen Sätzen auf (1) oder nicht (0; Dunning, 1993). Wenn sie jedoch häufig genug gemeinsam auftreten, wird diese dummy-kodierte Variable (0/1) mit dem log-transformierten χ^2-Wert multipliziert. Hier fand sich auch eine kontinuierlich skalierende Variante des Gesetzes der Frequenz wieder, die über die 0/1-kodierte Verwendung hinausgeht. Diese beiden Bestandteile fassten wir in einem einzigen Wert zusammen, der algorithmisch definierten Assoziationsstärke (AS; Hofmann et al., 2011), welche der zentrale Untersuchungsgegenstand dieser Habilitation werden wird. In Betrachtung des dritten Assoziationsgesetzes stellte sich jedoch die Frage, ob sich über die AS auch das Gesetz der Ähnlichkeit zwischen Wörtern abbilden lässt.

Die einfachste Hypothese, die man dabei testen könnte, wäre, dass die AS auch semantische Ähnlichkeiten abbildet. Aber wenn Wörter ausschließlich eine starke direkte AS aufweisen, dann finden sich diese beispielweise häufig in Idiomen wieder, wie zum Beispiel „Teufel“ und „Detail“; doch es finden sich auch andere solche Wortpaare, wie zum Beispiel „Kälte“ und „Hunger“ (vgl. Roelke, Franke et al., 2018). Sicher werden auch viele semantisch ähnliche Wörter häufig in denselben Sätzen verwendet, wie zum Beispiel „Hammer“ und „Sichel“. Dennoch existiert mindestens eine Form der Ähnlichkeit, die sich hierüber nicht sehr gut abbilden lässt. So werden relativ synonyme Wörter, wie zum Beispiel „Hochzeit” und „Heirat”, zwar selten im gleichen Satz verwendet, aber sie kommen in ähnlichen Satzkontexten vor (Rubenstein & Goodenough, 1965). Beide treten zum Beispiel mit den Wörtern *Thron*, *Witwe* oder *Geliebte* auf (vgl. Hofmann et al., 2011). Dieses Prinzip machte sich Rapp (2002) zu Nutze, um damit (partielle) Synonyme automatisch zu entdecken. Ob sich eine solche Definition der Bedeutungsähnlichkeit jedoch auch für psychologische Aufgaben nutzen lässt, wie zum Beispiel für die Vorhersage von semantischen Priming-Effekten, war zu Beginn dieser Habilitation noch unklar (Hutchison, 2003; Lucas, 2000).

2. Neurokognitive Modellierung semantischer Gedächtnisprozesse im Associative Read-Out Model (AROM): Überblick und Analysen

Publikation 1: Hofmann & Jacobs (2014)

Hofmann, M. J., & Jacobs, A. M. (2014). Interactive activation and competition models and semantic context: From behavioral to brain data. *Neuroscience and Biobehavioral Reviews*, *46*, 85–104.

In der ersten Publikation dieser Habilitation steht das IAM von McClelland und Rumelhart (1981) sowie seine Erweiterung um semantische und episodische Prozesse im Vordergrund (vgl. Abbildung 1A). Wir beginnen in Abschnitt 1 von Hofmann und Jacobs (2014) mit einer kurzen historischen Einordnung des IAMs in seinen unmittelbaren Entstehungskontext, führen ausgewählte Meilensteine seiner Entwicklung an und erläutern seinen Bezug zur modernen neurokognitiven Forschung (z. B. Price & Devlin, 2011).

In Abschnitt 2 fragen Hofmann und Jacobs (2014), wie Simulationsmodelle im Allgemeinen in der Psychologie entwickelt und validiert werden sollten. Dann beschreiben wir vier Evaluationskriterien für Theorie- und Modellentwicklung, die zum Beispiel zum Modellvergleich herangezogen werden können (vgl. Jacobs & Grainger, 1994): (i) Ein Modell sollte Daten möglichst genau beschreiben; (ii) es sollte einen möglichst allgemeingültigen Anwendungsbereich haben; (iii) es sollte einfach und damit möglichst leicht zu falsifizieren sein. Schließlich hat ein Modell, das mit möglichst wenigen Annahmen möglichst viele Phänomene möglichst genau erklärt, insbesondere dann einen hohen Erklärungswert (iv), wenn es die Vorhersage von Konsequenzen erlaubt, die vorher noch nicht bekannt waren (z. B. Grainger, O'Regan, Jacobs, & Segui, 1989).

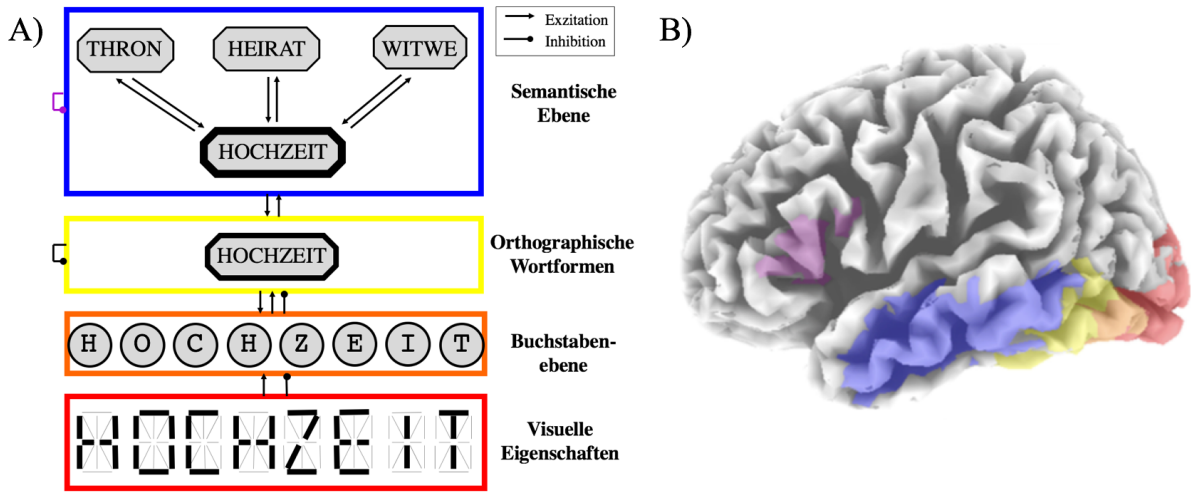

Abbildung 1. Um eine semantische Repräsentationsebene erweitertes IAM (Abb. 1A) und die entsprechenden Hirnregionen für Modell-Tests mit neurokognitiven Daten (Abb. 1B). **Panel A**: Die unteren drei Repräsentationsebenen entsprechen dem IAM. Ein Wort wird dem Modell als eine Reihe visueller Eigenschaften dargeboten. Die visuellen Eigenschaften, aus denen ein Buchstabe besteht, regen die entsprechenden Einheiten auf der Buchstabenebene an. Schließlich aktivieren die Buchstabeneinheiten die Einheiten auf der Ebene orthographischer Wortformen. Exzitation und Inhibition sind in dieser Abbildung jeweils als Pfeil mit spitzem und Punkt-Ende abgebildet. Die semantische Ebene entspricht der Modellerweiterung im AROM. Eine Worteinheit auf der semantischen Ebene wird durch das entsprechende Wort auf der orthographischen Ebene angeregt. Eine aktive Worteinheit auf der semantischen Ebene regt Aktivierung in assoziierten Wörtern an. Diese Exzitation wird an der AS skaliert (Hofmann & Jacobs, 2014). Da assoziierte Wörter Aktivierung zum präsentierten Wort zurücksenden, lässt sich zum Beispiel vorhersagen, dass Wörter, die viele assoziierte Wörter im Reizmaterial aufweisen, höhere Aktivierungen und damit höhere Ja-Antwortraten auslösen. **Panel B**: Hier werden die ungefähren Hirnregionen eingefärbt, die den farbigen Umrandungen der Repräsentationsebenen in Panel A entsprechen. Das visuelle Signal der Augen erreicht den visuellen Cortex (rot) etwa ab 50 ms nach Reizdarbietung. Diese Annahme lässt sich zum Beispiel durch Längeneffekte im okzipitalen Cortex (OC) und im Zeitbereich von 60 bis 125 ms stützen. In einem ähnlichen Zeitfenster zeigen sich auch semantische *Top-down*-Effekte durch die Vorhersagbarkeit eines Wortes aus dem Satzkontext, die sich im OC verorten lassen. Im IAM haben häufige Wörter höhere Ruheaktivierungen auf der Ebene orthographischer Wortformen. Deshalb können Frequenz-Effekte im linken fusiformen Gyrus (FFG, gelb) ab ungefähr 100 ms nach Reizdarbietung den Zugriff auf die orthographischen Wortformen anzeigen. Ab etwa 200 ms zeigen sich Effekte auf der N400-Komponente. Ab hier würden wir von einer Interaktion des *Bottom-up*-Signals des aktuellen Reizes mit dem semantischen Kontext ausgehen. Assoziative und semantische Effekte würden insbesondere in temporalen Cortices erwartet (blau). Semantische Repräsentationen sollten insbesondere im anterioren Temporallappen (ATL) repräsentiert werden. Semantische Selektion und Kompetition würde stattdessen insbesondere den linken inferioren Frontalgyrus betreffen (lila).

Abschnitt 3.1 bietet einen Überblick über ausgewählte Effekte in Verhaltens-, hirnelektrischen und funktionellen Bildgebungsstudien. Auf Grund dieses Überblicks schlagen wir eine Reihe von *model-to-data connections* vor (Jacobs & Grainger, 1994), mit Hilfe derer das IAM mit neurokognitiven Daten geprüft werden kann: Längeneffekte finden sich in ereigniskorrelierten Potentialen (EKPs) etwa ab 60 ms und im OC (z. B. Hauk & Pulvermüller, 2004; Schurz et al., 2010) und im Allgemeinen lösen kürzere Wörter schnellere Reaktionszeiten aus (Ziegler, Jacobs, & Klüppel, 2001; siehe aber New, Ferrand, Pallier, & Brysbaert, 2006; Oganian et al., 2016). Dies spiegelt die *bottom-up* getriebene, initiale Repräsentation der Stimuli auf der Ebene visueller Eigenschaften wider. Ab dieser Repräsentationsebene wirken auch Effekte der Wortsequenz: Wenn zum Beispiel der letzte Reiz in einer lexikalischen Entscheidungsaufgabe ein Wort war, können sich hier *top-down* getriebene Effekte ergeben, welche die Voraktivierung visueller Eigenschaften widerspiegeln (Kuchinke et al., 2011). Da die Wortfrequenz die Ruheaktivierungen der orthographischen Wortformen im IAM bestimmt, bilden antworterleichternde EKP-Wortfrequenzeffekte ab 100 ms und im FFG den Zugriff auf ein orthographisches Lexikon ab (z. B. Duncan, Pattamadilok, & Devlin, 2009; Hauk & Pulvermüller, 2004; Hofmann, Kuchinke, Tamm, Võ, & Jacobs, 2009; Kronbichler et al., 2004; Sereno, Rayner, & Posner, 1998). Effekte vor 200 ms und in dieser Region finden sich auch bei zumindest teilweisen orthographischen Überlappungen zwischen Prime- und Zielwort (z. B. Devlin, Jamison, Gonnerman, & Matthews, 2006; Huber, Curran, O‘Reilly, & Woroch, 2008), die sich durch fortdauernde Aktivierungen orthographischer Wortformen erklären lassen. Wenn ein Wort viele sehr ähnliche orthographische Wortformen aufweist, wie sich zum Beispiel an der Anzahl an Wörtern zeigen lässt, die sich durch genau einen Buchstaben vom Stimulus unterscheiden, dann wird er einfacher wiedererkannt (z. B. Grainger & Jacobs, 1996). Solche orthographischen Nachbarschafts-Effekte finden sich auf der N400-Komponente etwa ab 200 ms (Holcomb, Grainger, & O’Rourke, 2002; Kutas & Federmeier, 2011). Dass die Anzahl orthographischer Nachbarn bei Wörtern zu Effekten in temporalen Arealen führt (Binder et al., 2003), weist darauf hin, dass eine geringe orthographische Aktivierung kompensatorische Effekte in einer semantischen Ebene auslöst, die in IAMs zu diesem Zeitpunkt noch nicht abgebildet werden konnten (Coltheart et al., 2001; Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981; Perry et al., 2007).

In Abschnitt 3.2 greifen wir die Idee der *Conflict-monitoring*-Theorie auf, dass die Aktivierung vieler konkurrierender Einheiten zu Aktivierungen im *anterior cingulate cortex* (ACC) und zu

stärkeren Negativierungen auf der N2-Komponente führt (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Yeung, Botvinick, & Cohen, 2004). Da die simulierte Kompetition zwischen den orthographischen Einheiten in einem IAM mehr *Item-level*-Varianz an der N400-Komponente erklärt als die Anzahl orthographischer Nachbarn (Hofmann, Tamm, et al., 2008), und diese beiden Komponenten Ähnlichkeiten, wie zum Beispiel ein frontales Maximum und eine Quelle im ACC aufweisen, schlagen wir vor, die N400 als einen Spezialfall einer N2 zu betrachten. Da die N400 insbesondere auf semantische Manipulationen Effekte auslöst (Kutas & Federmeier, 2011), und sich N400-*item-level*-Varianz durch ein IAM nur für relativ bedeutungsfreie Nichtwort-Stimuli aufklären lässt (Hofmann, Tamm, et al., 2008; vgl. Braun et al., 2006), verdeutlicht dies die Notwendigkeit, diese Modelle um eine Ebene der Semantik zu erweitern. Klassische Ansätze der EKP-Forschung definieren semantische Erwartungen aus dem Satzkontext durch Lückentextergänzungswahrscheinlichkeiten (LTEW; z. B. Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Taylor, 1953). Hier lässt sich semantische Kompetition am einfachsten durch die Anzahl möglicher Ergänzungen operational definieren, die bei Wörtern mit einer geringen LTEW automatisch höher ist, da diese Wörter automatisch mehrere mögliche Ergänzungen nach sich ziehen (vgl. Kutas & Hillyard, 1984). Deshalb lassen sich höhere N400-Amplituden bei Wörtern mit geringer LTEW auch durch semantische Kompetition erklären. Die Frage, ob ein N400-Effekt besser durch die geringe Vorhersagbarkeit oder durch semantische Kompetition zu Stande kommt, ist auf Grund von LTEW nur schwer zu beantworten. Hier fehlen explizite komputationale Ansätze, die eine theoretische Trennung dieser beiden Prozesse ermöglichen (vgl. Thompson-Schill & Botvinick, 2006).

Schließlich wird in Abschnitt 4.1 die Frage gestellt, wie sich IAMs am besten um Bedeutungszusammenhänge zwischen Wörtern erweitern lassen, und wie sich der Geltungsbereich von IAMs hierdurch erweitern lässt. Für Priming-Aufgaben und beim Wiedererkennen studierter Wörter operationalisieren psychologische Studien semantische Bedeutungszusammenhänge typischerweise durch Vorexperimente mit der freien Assoziationsaufgabe (z. B. Lucas, 2000; Roediger & McDermott, 1995). Wir diskutieren Probleme eines solchen Ansatzes. Beispielsweise zeigten McKoon und Ratcliff (1992) Priming-Effekte für assoziierte Prime-Zielwort-Paare, die nicht in einer freien Assoziationsaufgabe genannt wurden, und schlugen vor, dass sich durch Kookkurrenzen auch

schwache Assoziationen feststellen lassen. Ein anderes Problem ist die Zirkularität, die entsteht, wenn man die menschliche Performanz in einer Aufgabe nutzt, um damit andere menschliche Leistungen zu erklären. Als einfache Lösung schlagen wir vor, zwei Wörter als assoziiert zu definieren, wenn sie signifikant häufiger gemeinsam in den Sätzen eines großen Korpus auftreten als deren Einzelauftretenshäufigkeiten erwarten ließen (Dunning, 1993; Quasthoff, Richter, & Biemann, 2006). An dieser Stelle werfen wir die Frage auf, ob die direkten Assoziationen zwischen Wörtern syntagmatische Relationen abbilden, und ob die indirekte assoziative Aktivierung über gemeinsame Assoziierte zweier Wörter sogenannte paradigmatische Relationen abbilden, wenn wir ein vollständiges Lexikon verwenden würden (vgl. Kapitel 6.1). Da diese Unterscheidung auch in den Kapiteln 4.1 und 4.4 dieser Habilitation weiterentwickelt wird (Hofmann et al., 2018; Roelke, Franke et al., 2018), sollen sie an dieser Stelle in dem Exkurs „Syntagmatische und paradigmatische Relationen im AROM“ zusammenfassend eingeführt werden.

Exkurs: Syntagmatische und paradigmatische Relationen im AROM

In seinem wegweisenden Buch „*Cours de linguistique générale*” schlug Ferdinand de Saussure (1959) zwei verschiedene Typen an Wortrelationen vor, die verschiedenen „*mentalen Aktivitäten*” entsprechen (de Saussure, 1959, p. 123[19]). Zum einen „*erhalten Wörter eine Relation durch die lineare Natur der Sprache, weil sie aneinander gekettet werden*“ (de Saussure, 1959, p. 123[20]). Er bezeichnet kurze idiomatische Äußerungen, wie „*Gott ist gut*” oder „*das Wetter ist schön*”, als Syntagma. In einem solchen Syntagma erhält ein Begriff seinen Wert, weil es zu den vorangegangenen oder folgenden Wörtern in Beziehung steht. Zum anderen können „*Wörter im Gedächtnis assoziiert sein, wenn sie etwas gemeinsam haben*” (De Saussure, 1959, p. 123[21]). Diese Verknüpfungen nennt De Saussure (1959) „*assoziative*

---

[19] „They correspond to two forms of our mental activity”.

[20] „words acquire relations based on the linear nature of language because they are chained together.”; „The syntagm is always composed of two or more consecutive units (e.g. French re-lire 're-read,' contre tous 'against everyone,' la vie humaine 'human life,' Dieu est bon 'God is good,' s'il fait beau temps, nous sortirons 'if theout,' etc.). In the syntagm a term acquires its value only because it stands in opposition to everything that precedes or follows it, or to both”.

[21] „Outside discourse, on the other hand, words acquire relations of a different kind. Those that have something in common are associated in the memory (…) Those formed outside discourse are not supported by linearity. Their seat is in the brain; they are a part of the inner storehouse that makes up the language of each speaker. They are associative relations.”

Relationen". Um begriffliche Unschärfen innerhalb dieser Habilitation zu vermeiden, sollen sie im Folgenden mit Reinhard Rapp (2002) als paradigmatische Relationen bezeichnet werden.

Während das direkte gemeinsame Auftreten in Sätzen (Dunning, 1993) als eine operationale Definition syntagmatischer Relationen betrachtet werden kann (vgl. die Definition der AS in Kapitel 1), nutzte Rapp (2002) die gemeinsamen Assoziierten von zwei Wörtern, um damit deren semantische Ähnlichkeit zu definieren. Die semantische Überlappung zwischen „rot" und „blau" lässt sich aus der Tatsache ableiten, dass beide zum Beispiel mit den Wörtern *Farbe*, *Blume* und *Kleidung* gemeinsam auftreten, die wir als semantische Eigenschaften betrachten (Stuellein et al., 2016). Wenn zwei Wörter viele gemeinsame Assoziierte aufweisen, aber nicht direkt gemeinsam in Sätzen auftreten, dann handelt es sich dabei zum Beispiel häufig um Synonyme (Rapp, 2002).

Stellen wir uns ein AROM mit einem vollständig implementierten Lexikon vor: Würden wir dem Modell zum Beispiel „blau" als Prime präsentieren, dann würde dieses Wort Aktivierungen in diesen gemeinsamen Assoziierten auslösen, zum Beispiel in *Farbe* und *Blume*. Diese wiederum würden das Wort „rot" anregen. Da das Zielwort „rot" demnach indirekte assoziative Aktivierungen vom Prime erhält, sollte es höhere Aktivierungen auf der semantischen Ebene erhalten und demnach leichter zu identifizieren sein (vgl. Kapitel 6.1).

Auf Grund der zeitlichen Dynamik im AROM lag die Hypothese nahe, dass zuerst direkte, eher syntagmatische Assoziationen wirken könnten: Da die Aktivierungen von einem rein paradigmatisch relatierten Prime erst über andere, gemeinsame Assoziierte fließen müssen, sollten die Effekte gemeinsamer Assoziierter eher bei einem längeren Intervall zwischen Prime- und Zielwort auftreten. Eine direkte Assoziation sollte jedoch früher wirken. Diese „*syntagmatic-first*" Hypothese ließ sich bei Erwachsenen später nicht bestätigen (Kapitel 4.4; Roelke, Franke et al., 2018; vgl. Friederici, 2002): Während bei einer kurzen *stimulus onset asynchrony* (SOA) von 200 ms direkte und gemeinsame Assoziationen Effekte auslösen, zeigen sich keine Effekte gemeinsamer Assoziierter bei einer sehr langen SOA von einer Sekunde. Da bei der langen SOA keine Priming-Effekte gemeinsamer Assoziierter mehr nachzuweisen sind, scheint hier eher „*paradigmatic first*" zu gelten. Im Entwicklungsverlauf hingegen lässt sich die „*syntagmatic-first*" Hypothese bestätigen: Während Kinder der zweiten Klasse bereits von direkten Assoziationen zwischen Prime- und Zielwort profitieren, nutzen

Kinder erst ab der vierten Klasse viele gemeinsame Assoziierte (Franke, Roelke, Radach, & Hofmann, 2017).

In Abschnitt 4.2 zeigen wir, wie sich die Anzahl assoziierter Wörter verwenden lässt, um damit menschliche Leistungen in einer episodischen Gedächtnisaufgabe vorherzusagen (Hofmann et al., 2011). In einer solchen Aufgabe lernen die Versuchspersonen Wörter in einer Studierphase und sollen diese von neuen, nicht-studierten Wörtern in einer Testphase unterscheiden. Wenn nicht-studierte Wörter eine hohe Anzahl Assoziierter in dieser episodischen Gedächtnisaufgabe aufweisen, dann führen stärkere assoziative Aktivierungen auf der semantischen Ebene zu Fehl-Erinnerungen bei in der Studierphase nicht gelernten Wörtern (Roediger & McDermott, 1995). Weist ein studiertes Wort hingegen viele Assoziierte auf, lässt sich eine Verbesserung der Gedächtnisleistung beim Wiedererkennen vorhersagen (z. B. Kimball, Smith, & Kahana, 2007). Diese Effekte können durch die dynamische Ko-Aktivierung assoziierter Wörter aus dem sprachlichen Kontext dieser Aufgabe simuliert werden (Hofmann et al., 2011; vgl. Cox & Shiffrin, 2017).

Abschnitt 4.3 bietet einen Beitrag zur Debatte über Ein- oder Zwei-Prozess-Modelle des episodischen Gedächtnisses. Während Yonelinas (1994) für studierte Wörter einen über die bloße Vertrautheit (*familiarity*) von nicht-studierten Wörtern hinausgehenden Prozess postuliert (*recollection*; vgl. Kuchinke, Fritzemeier, Hofmann, & Jacobs, 2013), beschreiben andere Messmodelle die Gedächtnissignalverteilungen von studierten Wörtern mit einer größeren Varianz bei studierten im Vergleich zu nicht-studierten Wörtern (z. B. Glanzer, Kim, Hilford, & Adams, 1999). An dieser Stelle versuchen wir, die in Hofmann et al. (2011) postulierte, algorithmische Erklärung für diese nicht falsifizierbare Messmodell-Annahme möglichst allgemeinverständlich in vier Schritten zu erklären (vgl. Appendix A, Hofmann & Jacobs, 2014, p. 100, für eine formale Beschreibung der im Folgenden präsentierten Logik). Wir benötigen die erste Messmodell-Annahme von Glanzer et al. (1999):

(i) **Studierte Wörter** weisen im Vergleich zu nicht-studierten Wörtern ein **höheres Gedächtnissignal** auf. Die Gedächtnissignale werden im AROM als positive Ruhe-Aktivierungen von Worteinheiten auf der semantischen Ebene dargestellt und sind demnach höher für studierte Wörter (vgl. die Wort-Frequenz-abhängigen Ruheaktivierungen auf der orthographischen Ebene im IAM, McClelland & Rumelhart, 1981).

Wenn diese Annahme in ein IAM implementiert wird, ergibt sich die folgende Beobachtung in der semantischen Ebene des AROM (Hofmann et al., 2011):

(ii) Wenn viele Wörter eine hohe Ruhe-Aktivierung aufweisen (i), resultiert daraus eine starke Kompetition und **Inhibition** zwischen den Wörtern, das heißt, nachdem alle exzitatorischen Eingangssignale hinzuaddiert wurden, erhalten alle Worteinheiten in der Summe ein inhibitorisches Netto-Input-Signal von den vielen anderen Worteinheiten mit hoher Aktivierung.

Die folgende Annahme ist im IAM bereits enthalten:

(iii) Eine **Nonlinearitätsannahme** des IAMs besagt, dass die in eine Repräsentation eingehenden exzitatorischen oder inhibitorischen Signale immer an der **aktuellen Aktivierung** der Repräsentation **skaliert** werden, wenn die Aktivierungsänderung berechnet wird (vgl. Grossberg, 1978). Wenn in der Summe ein exzitatorisches Signal vorliegt, dann wird eine Formel verwendet, die ähnlich zu der von Rescorla und Wagner (1972) ist (vgl. McClelland & Rumelhart, 1981, Formel 2, p. 380; Kapitel 1.1). Wenn in der Summe ein inhibitorisches Signal vorliegt (vgl. ii), dann findet Formel 3 aus McClelland und Rumelhart (1981, p. 381) Anwendung. Der inhibitorische Netto-Input wird mit der Differenz aus aktueller (Ruhe-)Aktivierung und minimaler Aktivierung multipliziert.

Hieraus ergibt sich folgende logische Konsequenz im Zyklus 1 des Modells:

(iv) Selbst wenn die Verteilungen inhibitorischer Netto-Input-Signale (ii) für studierte und nicht-studierte Wörter nun konstant wären, dann sorgen die höheren Gedächtnissignale studierter Wörter (i) durch Skalierung (iii) automatisch dafür, dass eine höhere Varianz des Gedächtnissignals entsteht.

Glanzer et al. (1999) schlugen ein Messmodell mit zwei unfalsifizierbaren Ad-hoc-Annahmen vor, die jeweils ein empirisches Phänomen beschreiben und damit keine Vorhersage erlauben: die Annahmen höhere Gedächtnissignale und höherer Gedächtnissignalvarianzen für studierte Wörter. In Hofmann und Jacobs (2014) zeigen wir, dass die erste Annahme, implementiert in eine Wortrepräsentationsebene eines IAMs, die zweite Annahme automatisch vorhersagt. Dies bietet eine Erklärung für die zweite Ad-hoc-Annahme des Messmodells von Glanzer und Kollegen (1999). Im Gegensatz zu dieser lassen sich aus den Erklärungsschritten ii-iv jedoch neue Vorhersagen ableiten, die falsifiziert werden können. Wenn zum Beispiel eine Variation

in der Lernzeit vorliegt (z. B. 1 vs. 3 Sekunden Präsentation der Wörter in der Studierphase), dann können wir gemäß (i) davon ausgehen, dass ein noch längeres Lernen zu einem noch größeren Gedächtnissignal führt. Da die Aktivierungsänderung immer an diesem Gedächtnissignal skaliert wird (iii), müsste ein längeres Lernen zu einer noch höheren Varianz der gelernten Wörter im Vergleich zu kürzer studierten Wörtern führen. Diese Hypothese lässt sich auf Basis der Literatur stützen (vgl. Glanzer et al., 1999): Je länger die Studierzeit, desto größer die relative Gedächtnissignalvarianz der studierten im Vergleich zu den nicht-studierten Wörtern. Die grundlegende IAM-Architektur kann folglich auch Phänomene erklären, für die es nicht geschaffen wurde, was ihren Erklärungswert erhöht. Die Annahme stärkerer Gedächtnissignale bei studierten Wörtern erklärt neben dem höheren Mittelwert der Ja-Antwortrate auch die höhere Gedächtnissignalvarianz ohne weitere Ad-hoc-Annahmen (Hofmann et al., 2011; Hofmann & Jacobs, 2014; vgl. Kapitel 5.1).

In Abschnitt 4.4 beschäftigen wir uns mit der Frage, in welchen Regionen des Gehirns semantische Effekte zu erwarten wären. Der Ansatz verkörperter Repräsentationen der Semantik würde vorhersagen, dass semantische Effekte nicht in einer einzigen Region zu verorten wären. Stattdessen werden semantische Informationen in den Hirnregionen repräsentiert, die auch die konkreten Objekte repräsentieren, welche die Begriffe bezeichnen (z. B. Pulvermüller & Fadiga, 2010). So würde ein Begriff, der sich auf den Arm bezieht, in den Subregionen des Motor-Cortex repräsentiert, der auch bei Bewegungen des Arms aktiv würde (z. B. Kiefer & Pulvermüller, 2012). Dennoch weisen Befunde aus Läsionsstudien darauf hin, dass auch temporale Areale und insbesondere der ATL semantische Informationen repräsentiert (Binder, Desai, Graves, & Conant, 2009; Patterson, Nestor, & Rogers, 2007) – Regionen, die wahrscheinlich auch eine neuronale Quelle der N400-Komponente darstellen (Lau, Phillips, & Poeppel, 2008). Darüber hinaus befindet sich tief im inneren der Temporallappen die Hippocampi, die insbesondere neue Assoziationen knüpfen (Kumaran & McClelland, 2012; vgl. Kapitel 6.2).

Eine weitere Steigerung des Erklärungswertes des AROMs erreichen wir in Abschnitt 4.5 durch eine Re-Analyse von Daten, die mit funktioneller Magnetresonanztomographie (fMRT) gemessen wurden (Forgács et al., 2012, $N = 40$). Eine hohe AS zwischen den Nomina von Komposita sagt erfolgreich eine Verringerung der hämodynamischen Antworten im linken inferioren Frontalgyrus (IFG) voraus, was sich am besten durch eine verringerte semantische Kompetition erklären lässt (Thomson-Schill & Botvinick, 2006; vgl. Kapitel 6.1). Dieser Effekt

ist so robust, dass er selbst bei einer vollen Bonferroni-Korrektur für über 90.000 statistisch als unabhängig postulierte Signifikanz-Tests signifikant ist. Schließlich blicken wir in Abschnitt 4.6 auf moderne Messmodelle der fMRT-Forschung und überlegen, wie die theoretisch postulierte Konnektivität verschiedener neurokognitiver Repräsentationsebenen in IAMs am besten mittels funktioneller Konnektivitätsanalysen geprüft werden kann (z. B. Friston, Harrison, Penny, 2003).

Abschnitt 5 von Hofmann und Jacobs (2014) adressiert die Frage, wie die semantische Kohäsionshypothese der emotionalen Wortverarbeitung am besten mit dem AROM getestet werden kann (Hofmann et al., 2011). Diese Hypothese besagt, dass ein großer Teil der Varianz in episodischen Gedächtnisaufgaben, der früher der emotionalen Valenz zugeschrieben wurde, sich eigentlich dadurch erklären lässt, dass affektiv geladene Wörter stärker assoziativ mit anderen Wörtern verknüpft sind (z. B. Maratos, Allan, & Rugg, 2000). Diese These lässt sich zum Teil durch eine positive Korrelation zwischen der Valenz der Wörter und der Anzahl Assoziierter stützen: Je positiver ein Wort von Versuchspersonen bewertet wird, desto mehr Assoziierte weist es in der *Berlin affective word list* auf (Võ et al., 2009). Dann präsentieren wir die Fehlerinnerungsergebnisse zweier Experimente beim Wiedererkennen nicht-studierter Wörter: Wenn sowohl die Anzahl assoziierter Wörter im Reizmaterial (hoch/niedrig), als auch die negative Valenz von Wörtern (negativ/neutral) experimentell manipuliert wird (N = 29), dann erhöht sowohl die Anzahl Assoziierter als auch die negative Valenz die Anzahl falscher Alarme. Wenn wir sowohl die Anzahl Assoziierter (hoch/niedrig) als auch die positive Valenz (positiv/neutral) in Betracht ziehen (N = 34), dann beobachten wir keinen Effekt negativer Valenz, sondern nur einen Effekt der Anzahl der Assoziierten im Reizmaterial. Dies zeigt, dass die Konfundierung zwischen positiver Valenz und der Anzahl assoziierter Wörter die Effekte positiver Valenz potentiell erklären kann. Es stellt sich die Frage, ob ein Modell, das solche assoziativen Effekte erklären kann, noch einen zusätzlichen Evaluationsmechanismus positiver Valenz braucht. Somit lässt sich der Geltungsbereich des AROMs auf die Effekte positiver Valenz erweitern und sein Erklärungswert weiter steigern.

3. Klassischer Ansatz der Lückentextergänzungswahrscheinlichkeiten (LTEW) zur Vorhersage hämodynamischer Antworten beim Satzlesen

Publikation 2: Hofmann et al. (2014)

Hofmann, M. J., Dambacher, M., Jacobs, A. M., Kliegl, R., Radach, R., Kuchinke, L., Plichta, M. M., Fallgatter, Herrmann, M. J. (2014). Occipital and orbitofrontal hemodynamics during naturally paced reading: An fNIRS study. *NeuroImage*, *94*, 193–202.

In Hofmann et al. (2014) untersuchen wir, wie die Vorhersagbarkeit eines Wortes aus dem Satzkontext die hämodynamischen Antworten im OC und orbitofrontalen Cortex (OFC) beeinflusst. Mit einer klassischen Operationalisierung der Wortvorhersagbarkeit auf Basis der LTEW testen wir hier, ob bei annähernd natürlichem Lesetempo *Top-down*-Effekte von semantischen Repräsentationen bis hin zur Repräsentationsebene visueller Eigenschaften wirken (vgl. Abb. 1). Aus der EKP-Literatur lässt sich die Vorhersage ableiten, dass frühe Vorhersagbarkeitseffekte nur dann zu Tage treten, wenn die Wörter hinreichend schnell nacheinander präsentiert werden (z. B. Dambacher et al., 2012; Dambacher, Rolfs, Göllner, Kliegl, & Jacobs, 2009). Würde man jedes Wort eines Satzes viel länger als beim natürlichen Lesen präsentieren, dann wären keine *Top-down*-Vorhersagen nötig, und die Reize könnten in einem reinen *bottom-up mode* in aller Ruhe entziffert werden (vgl. Bar et al., 2006).

Will man jedoch hämodynamische Antworten bei annähernd natürlichem Lesetempo untersuchen, dann würde die zeitliche Messgenauigkeit der goldenen Standard-Methode der hämodynamischen Antwort, das fMRT, diese Beobachtung sicher erschweren. Beim natürlichen Lesen wird ein Wort in der Regel etwa 200-250 ms fixiert, gefolgt von einer etwa 30 ms andauernden Blickbewegung, der Sakkade (Rayner, 1998, pp. 373, 375). Bei einer typischen zeitlichen Abtastrate beim fMRT von 0,5 Hz (z. B. Yarkoni, Speer, Balota, McAvoy, & Zacks, 2008), könnte die Versuchsperson etwa 7-9 Wörter während eines einzelnen Messzeitpunktes lesen. Damit die hämodynamischen Antworten auf einzelne Wörter besser voneinander separiert werden können, nutzen wir die Methode der funktionellen Nah-Infrarotspektroskopie (fNIRS) mit einer zeitlichen Auflösung von 10 Hz. Wir präsentieren die Sätze des Potsdam Satzcorpus 3 (PSC-3) in einem *rapid serial visual presentation* (RSVP)

Paradigma (Dambacher, 2010; Raymond, Shapiro, & Arnell, 1992). Um damit das natürliche Lesen möglichst gut anzunähern, präsentieren wir die Zielwörter im Satzkontext für 250 ms, gefolgt von einem 30 ms andauernden weißen Bildschirm – der jeweils typischen Dauer von Fixationszeit und Sakkade. Auch beim natürlichen Lesen wird während dieser Zeit kaum neue Information aufgenommen – ein Phänomen, das man mit dem Begriff *saccadic suppression* umschreibt (z. B. Rayner, 1998, p. 373).

Neben der Wortvorhersagbarkeit wird in dieser Studie auch der Faktor Wortfrequenz experimentell manipuliert (Dambacher et al., 2009, 2012). Die jeweiligen hoch- oder niederfrequenten Zielwörter werden in einen, in punkto Vorhersagbarkeit relativ neutralen Satz eingebettet (Dambacher et al., 2012). Die Manipulation der Wortvorhersagbarkeit wird über einen vorangestellten Kontext-Satz erreicht. Es existieren zwei verschiedene Kontext-Sätze, für die entweder das hochfrequente oder das niederfrequente Wort jeweils vorhersagbar ist – für den jeweils anderen Kontext-Satz sind diese Wörter kaum vorhersagbar. Eine vollständige, zweifaktorielle Manipulation wird über ein *Latin-square*-Design erreicht, so dass jeweils 7 der 28 Versuchspersonen jeweils eine Version dieser Sprachkontexte und Zielwörter zu sehen bekommen (vgl. Abb. 1, Hofmann et al., 2014). Für die ereigniskorrelierte Auswertung der fNIRS-Daten werden nicht nur die Präsentationszeitpunkte der vier Zielwortbedingungen, sondern auch alle nicht-interessierenden Ereignisse, wie zum Beispiel die Präsentationszeitpunkte des Kontext-Satzes oder der Nicht-Zielwörter, jeweils als Ereigniszeitpunkte definiert, bei denen jeweils eine hämodynamische Antwort ausgelöst wird. So können wir in einer Dekonvolutionsanalyse die als Störvariablen modellierten hämodynamischen Antworten aller nicht-interessierenden Ereignisse abziehen und damit die hämodynamischen Antworten der Zielwort-Effekte deskriptiv in ihrem Zeitverlauf abbilden (Abb. 6 & 7, Hofmann et al., 2014).

Die Ergebnisse bestätigen die Hypothese semantischer *Top-down*-Prozesse, die bis hin zur Ebene visueller Eigenschaften reichen: Eine erhöhte Ausschwemmung verbrauchten Blutes, das heißt desoxygenierten Hämoglobins, kann bei Wörtern mit einer geringen Wortvorhersehbarkeit im OC beobachtet werden. Die vereinfachende Annahme in IAMs einer nicht implementierten *Top-down*-Verknüpfung zwischen der Buchstaben- und visuellen Eigenschaftsebene müsste zumindest für die IAM-Simulation des Satzlesens mit einer nahezu natürlichen Leserate verworfen werden; ebenso wie die vereinfachende Annahme einer

fehlenden *Top-down*-Verknüpfung von der semantischen zur orthographischen Ebene (vgl. Abb. 1, Hofmann et al., 2011; vs. Abb. 1, Kapitel 2; Hofmann & Jacobs, 2014). Nur wenn vollständige *Bottom-up*- und *Top-down*-Interaktivität gewährleistet ist, wäre eine relativ einfache, modellbasierte Erklärung für solche Effekte, dass die visuellen Eigenschaften vorhersagbarer Wörter bereits vor der Reiz-Präsentation voraktiviert werden und deshalb weniger Energie für die Aktivierung der entsprechenden visuellen Eigenschaften im OC nötig ist.

Neben diesem Haupteffekt im OC, zeigt sich auch eine Interaktion der Wortfrequenz mit der Wortvorhersehbarkeit im OFC. Diese Interaktion lässt sich durch den signifikanten Paarvergleich innerhalb der selten in der Sprache auftretenden Wörter auflösen, dass die wenig vorhersagbaren Wörter signifikant mehr Sauerstoff verbrauchen als hoch vorhersagbare Wörter, was sich an der höheren Ausschwemmung desoxygenierten Hämoglobins für kaum vorhersagbare Wörter zeigt (Abb. 5, Hofmann et al., 2014). Wir schlagen die folgende Erklärung hierfür vor: Wenn ein Wort selten ist, wird es innerhalb von 250 ms Präsentationszeit noch nicht vollständig auf Grund der *Bottom-up*-Informationen identifiziert (Bar et al., 2006). Wenn dieses darüber hinaus auch nicht auf Grund des Satzkontextes vorhersagbar ist, dann wird ein Erwartungsverletzungs-Signal im OFC ausgelöst, wie es von Nobre, Coull, Frith und Mesulam (1999) vorgeschlagen wurde. Doch auch der gegenteilige Effekt ist aus der Literatur bekannt: Bar und Kollegen (2006) schlugen vor, dass der OFC aktiv wird, wenn die *top-down* Vorhersagen den eingehenden *bottom-up* Informationen entsprechen. Diese Vorhersage lässt sich mit den vorliegenden Daten innerhalb der hochfrequenten Wörter deskriptiv stützen: Wir finden hier eine nicht-signifikante Tendenz, bei hochvorhersagbaren Wörtern mehr Energie im OFC zu verbrauchen (Abb. 7, Hofmann et al., 2014). Wenn ein Wort häufig in der Sprache auftritt, ist es auf Grund von *Bottom-up*-Informationen innerhalb der 250 ms Reizpräsentation eher zu identifizieren. Wenn dieses hochfrequente Wort zusätzlich auf Grund „semantischer" *Top-down*-Informationen vorhersagbar ist, dann findet sich eine Tendenz für Aktivierung im OFC (vgl. Bar et al., 2006).

Dass sich ein solcher OFC-Interaktionseffekt durch die innerhalb einer vorgegebenen Zeit aufgenommenen Informationen in diesem nur leseähnlichen RSVP-Paradigma erklären lässt, konnten wir in einer Folgestudie beim natürlichen Lesen bestätigen (Roelke, Hofmann, & Radach, 2018). Als wir die zum Zeitpunkt einer Fixation ausgelöste hämodynamische Antwort durch gleichzeitige Blickbewegungsmessung und fNIRS untersuchten, fanden wir keine

Effekte mehr im OFC. Stattdessen fand sich ein ähnlicher Interaktionseffekt für die rechts von der Fixation im parafovealen Sichtfeld befindlichen Wörter retinotop im linken OC (Roelke, Hofmann et al., 2018; vgl. z. B. Larsson & Heeger, 2006; Radach, Inhoff, Glover, & Vorstius, 2013). Diesen Konferenzbeitrag werden wir in Kürze zur Veröffentlichung einreichen, um die Methode der fixationsrelatierten Nah-Infrarotspektroskopie (frNIRS) in die Leseforschung einzuführen (vgl. z. B. Marsman, Renken, Velichkovsky, Hooymans, & Cornelissen, 2002; Hofmann, Hermann, et al., 2008; Hutzler et al., 2007).

4. Algorithmische Modelle zur Vorhersage von Verhaltens- und neurokognitiven Daten

4.1 Assoziative Urteile bei Wortpaaren

Publikation 3: Hofmann et al. (2018)

Hofmann, M. J., Biemann, C., Westbury, C. F., Murusidze, M., Conrad, M., & Jacobs, A. M. (2018). Simple co-occurrence statistics reproducibly predict association ratings. *Cognitive Science*, *42*, 2287–2312.

In dieser Untersuchung testen wir die Hypothese, dass die in Kapitel 1 eingeführte AS reproduzierbar Varianz an Assoziations-*ratings* von Wortpaaren auf einer siebenstufigen *Rating*-Skala vorhersagt. Mittelt man diese Urteile über Items, das heißt Wortpaare, so klärt die AS etwa 40 % der *Item-Level*-Varianz in den drei durchgeführten Studien auf (je N ~ 30 Versuchspersonen). Diese Varianzaufklärung entspricht ungefähr der Modell-Performanz der erfolgreichsten IAMs (z. B. Perry et al., 2007), die jedoch noch nicht über eine implementierte semantische Repräsentationsebene verfügten. Für die Analysen mit multiplen Prädiktoren verwenden wir *linear mixed effects models*, die sich in den letzten Jahren immer mehr zum Standard psycholinguistischer Forschung entwickelten (z. B. Baayen, Davidson, & Bates, 2008). Sie erlauben es, getrennte Item- und Versuchspersonen-Analysen durch ein non-hierarchisches Messmodell zu ersetzen, so dass alle systematisch variierenden Prädiktoren Varianz auf der Datenebene der Einzeldurchgänge aufklären (*fixed effects*). Darüber hinaus dienen zufällig variierende, für jedes Item und jede Versuchsperson einzeln abgeschätzte Achsenabschnitte dazu, Fehlervarianz auf Grund zufälliger Variationen in Versuchspersonen und Items zu absorbieren, um somit eine bessere Generalisierbarkeit zu erreichen (*random effects*).

Zum einen nutzen wir diese Messmodelle, um eine Reihe geplanter Analysen durchzuführen. Für diese Analysen wurden die Reize so gewählt, dass die geplanten Prädiktor-Variablen relativ geringe Korrelationen aufweisen ($r \leq 0.3$). Neben der AS zeigt auch die Interaktion der emotionalen Valenz der beiden Wörter in allen drei Studien einen Effekt, nicht jedoch die Häufigkeit, das *arousal* oder die *imageability* der Wörter (Võ et al., 2009).

Zum anderen werden die assoziativen Urteile der drei Studien genutzt, um damit umfassendere, explorative Analysen mit einer Reihe alternativer, in der Computerlinguistik gängiger Assoziationsmaße durchzuführen. Da die Prädiktoren hier zum Teil deutlich höhere Korrelationen aufweisen und damit das Problem der Multikollinearität erheblich erscheint, identifizieren wir zunächst hochgradig korrelierte Prädiktoren mit einer agglomerativen hierarchischen Clusteranalyse (z. B. Montefinese, Ambrosini, Fairfield, & Mammarella, 2014). Dann untersuchen wir die Prädiktoren in kompetitiven, Cluster-weisen Analysen. Reproduzierbar über alle drei Studien konnten nur die AS und das *Skip-gram*-Modell Varianz aufklären – ein sehr erfolgreicher maschineller Lernalgorithmus, der im Prinzip auf dem berühmten Artikel „*Finding Structure in Time*" von Jeffrey Elman (1990) basiert. Da solche rekurrenten neuronalen Netzwerkmodelle auch an anderen Stellen dieser Habilitation eine Rolle spielen, möchte ich sie im folgenden Exkurs zusammenfassend erläutert (vgl. Kapitel 1).

Exkurs: Rekurrente neuronale Netzwerkmodelle

Elmans (1990) Modell postulierte eine Input- und eine Output-Ebene, die aus einem Vektor aus Nullen und Einsen bestand (vgl. Abb. 2, Elman, 1990). Jedes Wort im „Lexikon" des Modells entsprach einem Wert in diesem Vektor. Diesem Modell wurde jeweils ein aktuelles Wort auf der Input-Ebene präsentiert. Dann wurde das Modell darauf trainiert, das jeweils nächste Wort im Satz durch die Aktivierungen der Output-Ebene vorherzusagen. Für die Input-Ebene und das Trainings-Signal der Output-Ebene war der Wert des jeweiligen Wortes 1, wenn es präsentiert wurde. Alle anderen Wörter hatten den Wert 0. Diese beiden Ebenen waren durch eine Ebene aus *hidden units* verknüpft, die trainierbare Verbindungen zur Input- und Output-Ebene aufweisen. Sobald man das jeweils nächste Wort präsentiert, wurde der *Hidden-layer*-Zustand über nicht-trainierbare, sogenannte rekurrente Verknüpfungen, in eine Kontext-Ebene kopiert. Somit bildete diese Kontext-Ebene den Zustand der *hidden units* zum jeweils letzten Zeitpunkt ab. Diese Kontext-Ebene wiederum war über trainierbare Verknüpfungen mit der Ebene der *hidden units* verbunden, welche die Bedeutung des aktuellen Wortes abbilden. Da die Wortrepräsentationen der Vergangenheit die aktuellen *Hidden-unit*-Repräsentationen aktivieren, sprechen wir hier von rekurrenten neuronalen Netzwerken (RNNs). Durch die Kontext-Ebene erhielt das Modell sozusagen ein Kurzzeitgedächtnis, das die

Wortrepräsentationen im *hidden layer* beeinflusst (Mikolov, Chen et al., 2013, p. 13). Zu Beginn der Simulationen wurden alle trainierbaren Verknüpfungen zufällig initiiert (vgl. Abb. 2, Elman, 1990). Bei der Präsentation eines jeden Wortes wurde die Modellvorhersage für das nächste Wort auf der Output-Ebene mit dem tatsächlich als nächstes präsentierten Wort verglichen. Je stärker diese Abweichung, dieses sogenannte Delta, desto stärker wurden die Verbindungsgewichte verändert. So konnte Elmans (1990) Modell mit einer einfachen Delta-Regel darauf trainiert werden, das jeweils nächste Wort im Satz vorherzusagen. Elman (1990) zeigte, dass sich durch diesen Lernmechanismus zunächst syntaktische (Nomen, Verb), dann aber auch semantische Relationen zwischen Wörtern (belebt, unbelebt) als Distanz zwischen den Aktivierungen im Vektorraum der *hidden units* abbilden lassen (vgl. Abb. 7, Elman, 1990). Je länger das Modell trainiert wurde, desto feinere „Subkategorien" differenzieren sich in diesem Modell aus (z. B. Mensch, Tier innerhalb belebter Objekte).

Elman (1990, p. 194f) verwendete für seine exemplarischen Simulationen zum Beispiel ein Lexikon aus etwa 30 verschiedenen lexikalischen Einheiten. Diese lexikalischen Einheiten lassen sich insgesamt 12 exemplarischen, semantischen Kategorien zuordnen, wie zum Beispiel menschliche Nomen (Frau/Mann) oder Wahrnehmungsverben (riechen/sehen). Mit Hilfe von Regeln, die typische Aussagen abbilden, generierte er hieraus zum Beispiel 10.000 Sätze, um damit ein exemplarisches Trainings-Korpus zu erhalten. Mit der Verfügbarkeit immer größerer Rechenkapazitäten wurde diese grundlegende Modellarchitektur später für das Training an großen Korpora weiterentwickelt. Ein solches Modell werden wir an vielen Millionen Sätzen in Kapitel 4.2 trainieren (Hofmann et al., 2017; Mikolov, 2012). Diese konkreten Modelle werden wir in dieser Synopsis als *simple recurrent network* (SRN) bezeichnen. Als weiterer Meilenstein in der Entwicklung von RNNs können die *Word2vec*-Modelle von Mikolov, Chen et al. (2013) betrachtet werden. Im Gegensatz zu SRNs, die darauf trainiert wurden, durch die vorangegangenen Wörter das jeweils nächste Wort vorherzusagen (Mikolov, 2012), wurden *Word2vec*-Modelle drauf trainiert, entweder durch den umliegenden sprachlichen Kontext ein Zielwort vorherzusagen oder umgekehrt (Mikolov, Chen et al., 2013): Im *Skip-gram*-Modell werden die *hidden units* trainiert, die beiden links und rechts im Satz auftretenden Wörter vorherzusagen. Im *continuous bag-of-words* (CBOW) Modell sagen diese umliegenden Wörter das Zielwort vorher. Diese beiden Modelle verwenden wir in Kapitel 4.1.

Die AS und das *Skip-gram*-Modell klären im ersten Cluster reproduzierbar Varianz an den assoziativen *ratings* über alle drei Studien auf. In diesem Cluster konnte sich weder das CBOW-Modell, noch die Anzahl gemeinsamer Assoziierter oder die aus der Informationstheorie abgeleitete (*positive*) *pointwise mutual information* – im Wesentlichen die gemeinsame Auftretenswahrscheinlichkeit dividiert durch das Produkt der einzelnen Auftretenswahrscheinlichkeiten der Wörter (Bouma, 2009) – reproduzierbar Varianz aufklären (vgl. Kapitel 1); ebenso wie die anderen Prädiktoren des zweiten Clusters (Worthäufigkeit und *arousal*), des dritten Clusters (*emotional valence* und *imageability*) oder des vierten Clusters, das nur aus einer Variable besteht: der Anzahl der Sätze, in denen die beiden Wörter gemeinsam auftreten. Für das Training der algorithmischen Modelle in Kapitel 4.1 wurde ein 43 Millionen Sätze umfassendes Korpus verwendet (Quasthoff et al., 2006).

Diese Ergebnisse zeigen, dass ein syntagmatischer, das heißt auf frequenzgewichteter Langzeit-Kontiguität basierender Ansatz, wie die AS, zusammen mit einem wahrscheinlich paradigmatische Relationen abbildenden Ansatz der semantischen Ähnlichkeit (Frank & Willems, 2017), zum Beispiel das *Skip-gram*-Modell, die einzigen Prädiktoren von Assoziations-*ratings* sind, die einen reproduzierbaren Varianzanteil aufklären. Die emotionale Valenz, die in den geplanten Analysen stets Varianz aufklärt, kann in den explorativen Analysen keinen reproduzierbaren Varianzanteil erklären. Das heißt, sie wird wahrscheinlich von der im *Skip-gram*-Modell latent enthaltenen semantischen Struktur als Prädiktor absorbiert (Hollis, Westbury, & Lefsrud, 2017; Westbury et al. 2015): Als Erklärung schlagen wir vor, dass die emotionale Valenz eine grundlegende „semantische Dimension" darstellt, die den „semantischen Raum" mit konstituiert. Deshalb kann sie auch in solchen Modellen des semantischen Gedächtnisses abgebildet werden (Jacobs et al., 2015; vgl. z. B. Osgood, Suci & Tannenbaum, 1957).

Somit lässt sich nicht nur erklären, warum man die Valenz sehr gut aus semantischen Modellen berechnen kann (z. B. Westbury et al., 2013), sondern auch andere Phänomene: Beispielsweise konnten die Mittelwerte der komputational errechneten Valenzen der Wörter eines Satzes zur Vorhersage der Valenz von Sätzen eingesetzt werden (Überblick bei Jacobs, Hofmann, & Kinder, 2016). Eine zweite Evidenz, die dafürsprach, dass emotionale Informationen aus semantischen Netzwerken extrahiert werden können, ergab sich aus einer *forced-choice* Valenzentscheidung: Hier wurden neutrale Wörter so beurteilt wie die Valenz direkt

assoziierter Wörter (Kuhlmann, Hofmann, Briesemeister & Jacobs, 2016). Schließlich führten inkongruente Valenzen der Nomina eines Kompositums zu Aktivierungen im IFG: Dies ließ sich durch Kompetition zwischen positiv und negativ konnotierten Bedeutungsfeldern erklären, die in „semantischen“ Regionen verarbeitet werden (Kuhlmann, Hofmann, & Jacobs, 2017).

Ein wesentlicher inhaltlicher Beitrag im Diskussionsteil von Hofmann et al. (2018) besteht darin zu fragen, welche Elemente computerlinguistischer Sprachmodelle welche spezifischen kognitiven Mechanismen im Gedächtnis abbilden. Wir schlagen vor, dass das Korpus, auf Basis dessen ein computerlinguistisches Modell trainiert wird, die „Erfahrungsgrundlage“ darstellt, auf deren Basis eine menschliche LZG-Struktur entsteht (vgl. Kapitel 6.3). Der spezifische computerlinguistische Algorithmus, zum Beispiel das *Skip-gram*-Modell, das am Korpus trainiert wird, bildet einen Mechanismus der Gedächtniskonsolidierung ab. Dieses Modell bietet eine Schätzung, wie genau aus der Erfahrungsgrundlage eine LZG-Struktur gebildet wird. Genauso wie die Versuchsperson im Experiment nutzt das *Skip-Gram*-Modell diese semantische Struktur, um damit Vorhersagen zu machen: Die Cosinus-Distanz im vieldimensionalen Vektorraum wird zwischen den austrainierten *Hidden-unit*-Repräsentationen der Wörter berechnet (Mikolov, Chen et al., 2013). Dieser Berechnungsschritt entspricht dem Abschätzen der semantischen Distanz der Versuchspersonen im LZG, während sie die assoziativen Urteile fällen.

Schließlich diskutieren wir den einfachsten Prädiktor, der reproduzierbar Varianz in allen Analysen aufklärt, die AS, die aus dem überzufällig häufig gemeinsamen Auftreten von Wörtern in Sätzen gewonnen wurde. Wie in Kapitel 1 aus der Geschichte der Assoziationsforschung hergeleitet, entspricht die AS einer formalen Definition der beiden Assoziationsgesetze der Kontiguität und der Frequenz aus aristotelischer Tradition (McKeon, 1941). Das Gesetz der Kontiguität besagt, dass die Erfahrung oder die Erinnerung eines Objektes die Erinnerung an Dinge auslöst, die ursprünglich mit diesem Objekt zusammen erlebt wurden. Das Gesetz der Frequenz besagt, dass je häufiger diese Dinge gemeinsam erlebt wurden, desto wahrscheinlicher wird es sein, dass die Erfahrung oder die Erinnerung des einen die Erinnerung an das andere stimuliert (Olson & Hergenhahn, 2017). Die resultierende AS kann als das Ergebnis eines einfachen, frequenzspezifischen Kontiguitäts-Lernmechanismus betrachtet werden, der am Beispiel eines Korpus trainiert wurde. Theoretisch bildet die Anzahl gemeinsamer Assoziierter zweier Wörter die semantische Überlappung dieser beiden Wörter ab (vgl. Rapp, 2002; Stuellein, Radach, Jacobs, & Hofmann, 2016; vgl. Kapitel 1 und 4.3). Für

diese, in Anlehnung an De Saussure (1959) als „paradigmatisch" zu bezeichnende Relation zwischen zwei Wörtern (z. B. Rapp, 2002), erlaubt jedoch das *Skip-gram*-Modell eine bessere Vorhersage als die einfache Anzahl gemeinsamer Assoziierter. Das *Skip-gram*-Modell benötigt dazu jedoch eine deutlich höhere Anzahl an Parametern. Diese Annahmen sind deutlich weniger gut verstanden als das SRN (Elman, 1990) oder die einfache Anzahl gemeinsamer Assoziierter (vgl. Levy, Goldberg, & Dagan, 2015). Eine wahrscheinliche Erklärung für die Überlegenheit des *Skip-gram*-Modells wäre, dass es nicht nur die gemeinsamen Assoziierten nutzt, sondern dass deren *hidden units* hierzu auch inhaltlich ähnliche Wörter berücksichtigen (Hofmann et al., 2018, p. 2306).

4.2 LTEW, Ereigniskorrelierte Potentiale (EKPs) und Blickbewegungen beim Satzlesen

Publikation 4: Hofmann et al. (2017)

Hofmann, M. J., Biemann, C., & Remus, S. (2017). Benchmarking n-grams, topic models and recurrent neural networks by cloze completions, EEGs and eye movements. In B. Sharp, F. Sedes, & W. Lubaszewsk (Eds.), *Cognitive Approach to Natural Language Processing* (pp. 197–215). London, UK: ISTE Press Ltd, Elsevier.

In Hofmann et al. (2017) testen wir, ob sich drei Standard-Sprachmodelle der Computerlinguistik dazu verwenden lassen, die LTEW, N400-Amplituden und Einzelfixationsdauern des PSC-1 vorherzusagen (Dambacher, 2010; Dambacher & Kliegl, 2007; Kliegl, Grabner, Rolfs, Engbert, 2004). Dabei beschreibt der Begriff der Einzelfixationsdauer, der *single-fixation duration* (SFD), das erfolgreiche Erkennen eines Wortes auf einen Blick, das heißt, es ist keine weitere Fixation nötig, um dieses Wort erfolgreich zu erkennen (z. B. Inhoff & Radach, 1998). Das erste Modell, mit dem wir diese Daten erklären, ist ein *N-gram*-Modell (Kneser & Ney, 1995), das positionsspezifische, bedingte Wahrscheinlichkeit abbildet, dass ein Wort auftritt, gegeben die vorausgehenden vier Wörter im Satz. Bei diesem 5-gram Modell handelt es sich, ähnlich wie die AS, um ein einfaches Maß frequenzgewichteter Kontiguität. Im Gegensatz zur AS bildet dieses Maß jedoch positionsspezifische Kontiguitäten zu den vorausgehenden Wörtern ab (vgl. Ebbinghaus, 1885, p. 124).

Bei der LSA wird das gemeinsame Auftreten von Wörtern in Dokumenten verwendet, um semantische Dimensionen mittels *single-value decomposition* zu berechnen (z. B. Landauer & Dumais, 1997). Ein alternativer, rein probabilistischer Ansatz ersetzt diese Methode der Dimensionsreduktion durch die LDA (Blei et al., 2003). Die resultierenden *Topic*-Modelle machen in psychologischen Aufgaben häufig bessere Vorhersagen als die LSA (Griffiths et al., 2007) und wurden deshalb auch in der aktuellen Studie als zweites Sprachmodell verwendet. Das *Topic*-Modell nutzt einen *Gibbs-sampling*-Algorithmus, um die Wahrscheinlichkeiten eines jeden Wortes w zum Dokument d zu gehören (Biemann, Remus, & Hofmann, 2015; Phan & Nguyen, 2007) durch die Summe der Wahrscheinlichkeiten über in unserem Fall 200 *topics* z abzuschätzen: $p(w|d) = \sum p(w|z) * p(z|d)$. Somit wird beim Training des Modells die empirisch

beobachtbare Wahrscheinlichkeit, dass ein Wort auftritt, gegeben ein Dokument, durch eine latente Verteilung der Zugehörigkeiten zu den *topics* approximiert. Die Zugehörigkeit zu jeweils einem *topic* z besteht aus dem Produkt von $p(w|z)$ und $p(z|d)$: der Wahrscheinlichkeit, dass ein Wort auftritt, gegeben das jeweilige *topic*; mal der Wahrscheinlichkeit, dass dieses *topic* auftritt, gegeben das jeweilige Dokument. Nachdem das Modell in der Phase der Gedächtniskonsolidierung trainiert wurde, liegen Wort-*topic* Matrizen vor, welche die Zugehörigkeit der Wörter zu den $N = 200$ *topics* abbilden. Ebenso stehen *Topic*-Dokument-Matrizen zur Verfügung, welche die Dokumente aufgrund der Ähnlichkeiten der *topics* klassifizieren. Nachdem dieses Training abgeschlossen ist, berechnen wir zur Simulation der Phase des Gedächtnisabrufs die Wahrscheinlichkeit, dass ein Wort auftritt, gegeben die vorangegangenen Wörter im Satz. Vereinfacht gesprochen, wird das d in der *Topic*-Grundgleichung durch die Summe der Wahrscheinlichkeiten der vorangegangenen Wörter ersetzt. Wir berechnen die *Topics*-Übereinstimmungen des aktuellen Wortes mit den vorangegangenen Wörtern im Satz. Da die semantische Distanz hier auf Grund der semantischen Relationen innerhalb eines Dokumentes errechnet wurde, sprechen wir hier von *Long-range*-Semantik. Sie sollte abbilden, welchen Effekt über längere Zeiträume als den Satz integrierte semantische Informationen auf den aktuellen Satzkontext hat.

Im Gegensatz dazu werden semantische Strukturen beim *N-gram*-Modell aus dem Satzkontext berechnet: Wir können hier von einem *Short-range*-Semantik-Modell sprechen. Als zweites *Short-range*-Semantik-Modell verwenden wir moderne SRNs, wie wir sie im Exkurs zu RNNs in Kapitel 4.1 eingeführt haben (Mikolov, 2012; vgl. Elman, 1990).

Die *N-gram*-, *Topic*- und SRN-Modelle werden an drei verschiedenen Korpora trainiert, um damit Unterschiede in der Erfahrungsgrundlage für potentielle LZG-Strukturen abzubilden. Wir verwenden ein *News*-Korpus, das aus 540 Millionen Wortformen, 30 Millionen Sätzen und 3,4 Millionen Dokumenten besteht. Letztere entsprechen Artikeln, die primär deutschen Online-Zeitschriften entnommen sind. Das zweite Korpus ist ein Wikipedia-Korpus, das 180 Millionen Wortformen, 7,7 Millionen Sätze und etwa 114.000 Artikel umfasst. Als drittes Korpus verwenden wir ein *Subtitles*-Korpus, das aus den Untertiteln von Filmen besteht und damit gesprochene Sprache abbildet (Brysbaert et al., 2011). Es enthält 54 Millionen Wortformen, 7,3 Millionen Äußerungen und 7.420 Filme und Fernsehserien, die für das Dokument-*level*-Training der *Topic*-Modelle verwendet wurden. Dieses Korpus ist deutlich

kleiner als die anderen Korpora. Brysbaert und Kollegen (2011) konnten jedoch zeigen, dass die hieraus gewonnenen Wortfrequenzmaße größere Korrelationen mit lexikalischen Entscheidungs- und Benennungslatenzen aufweisen als andere Korpora. Als mögliche Erklärung schlugen sie vor, dass die Sprache im Fernsehen repräsentativer für den alltäglichen Sprachgebrauch ist.

Um zu testen, wie viel Varianz die Sprachmodelle reproduzierbar vorhersagen können, teilen wir die von 272 Versuchspersonen gewonnenen empirischen LTEW für die 1138 Wörter des PSC-1 zufällig in zwei Teilstichproben zu je 569 Wörtern auf (Kliegl et al., 2004). Die beste *Item-level*-Varianzaufklärung bietet ein *N-gram*-Modell, das auf dem *News*-Korpus trainiert wurde. Zusammen mit den Baseline-Prädiktoren Wortfrequenz und der Position im Satz konnte es jeweils 46 % und 49 % der Varianz der LTEW an den beiden Teilstichproben aufklären. Diese Vorhersage wurde durch die Hinzunahme der anderen beiden Sprachmodelle kaum verbessert (47 % und 49 %). Insgesamt zeigt sich, dass man mit *N-gram*- und SRN-Modellen und der Baseline auf allen Trainings-Korpora reproduzierbar im Bereich von 42 % bis 49 % Varianz an den LTEW aufklären kann (vgl. Tabelle 10.2, Hofmann et al., 2017).

Für die Analyse von N400-Amplituden und SFDs lagen uns über Items gemittelte Daten für 343 Nomen aus Dambacher und Kliegl (2007) vor (jeweils $N = 48$ und $N = 125$), die wir in diesen Analysen auf Grund der geringen Stichprobengröße nicht in zwei Teilstichproben aufteilen. Wie für die LTEW, zeigt sich auch für die N400-Amplitude, dass das auf Basis des *News*-Korpus trainierte *N-gram*-Modell die beste Varianzaufklärung aufweist (16 %). Diese kann durch die Hinzunahme der anderen Sprachmodelle kaum verbessert werden (17 %). Die Hinzunahme der empirischen LTEW als Prädiktor bringt jedoch einen deutlichen Gewinn an Varianz (23 %). Wenn das *N-gram*- und das SRN-Modell an anderen Korpora trainiert wurden, können sie etwa 11 % bis 15 % der N400-Amplituden-Varianz aufklären.

Ein etwas anderes Ergebnismuster findet sich für die SFD (Dambacher & Kliegl, 2007). Hier zeigt das SRN mit Baseline die beste Varianzaufklärung von 28 %, wenn es auf dem *Subtitles*-Korpus trainiert wurde, gefolgt vom Training auf dem *News*-Korpus mit 27 %. Das *N-gram*-Modell klärt auf diesen Korpora jeweils 24 % und 23 % auf. Während wir durch Hinzunahme aller Sprachmodelle im *Subtitles*-Korpus bis zu 31 % Varianz erklären, kann durch die Hinzunahme der LTEW insgesamt bis zu 33 % an der SFD-Varianz aufgeklärt werden: Der

klassische Ansatz der LTEW scheint auch hier zumindest deskriptiv die SFD-Vorhersage zu verbessern.

An den SFDs klärt ein rein „klassisches" Regressionsmodell, bestehend aus den LTEW und der Baseline, alleine etwa 18 % der Varianz auf (vgl. z. B. Kliegl et al., 2004; Reichle, Rayner, & Pollatsek, 2003). Wenn wir dieses Regressionsmodell mit einem relativ konservativen Fisher-Yates *z*-Test mit dem Regressionsmodell vergleichen, das alle drei *Subtitles*-basierten Sprachmodelle und die Baseline enthält, dann zeigt sich ein signifikanter Unterschied. Die Sprachmodelle zusammen erlauben eine bessere Vorhersage als das „klassische" Modell (vgl. Abb. 10.3, Hofmann et al., 2017). Hier bleibt jedoch kritisch anzumerken, dass bei einer hohen Anzahl Prädiktoren die Wahrscheinlichkeit steigt, Fehlervarianz mit dem Messmodell abzubilden. Dennoch halten wir dieses Ergebnis, das auf einem relativ wenig sensitiven Fisher-Yates *z*-Test basiert, für beachtlich. Derzeit arbeiten wir daran, diese Studie mit wesentlich sensitiveren nonlinearen Messmodellen auf der Ebene von Einzeldurchgängen zu replizieren und zu erweitern (vgl. Kapitel 4.1; Wood, 2011).

4.3 EKPs in einer episodischen Gedächtnisaufgabe

Publikation 5: Stuellein et al. (2016)

Stuellein, N., Radach, R. R., Jacobs, A. M., & **Hofmann**, M. J. (2016). No one way ticket from orthography to semantics in recognition memory: N400 and P200 effects of associations. *Brain Research*, *1639*, 88–98.

Wenn man das AROM in seiner ersten Fassung (Hofmann et al., 2011) aufmerksam mit Abbildung 1 vergleicht (Hofmann & Jacobs, 2014), dann lässt sich als inhaltlicher Unterschied die *Top-down*-Verknüpfung zwischen der semantischen und der orthographischen Ebene feststellen (vgl. Abb. 1). Um die Simulationen in Hofmann et al. (2011) zu vereinfachen, einen freien Parameter einzusparen und damit im Sinne des *nested incremental modeling* auch ein von der semantischen Ebene unbeeinflusstes IAM zu behalten (Jacobs & Grainger, 1994), wurde in Hofmann et al. (2011) die vereinfachende Annahme gemacht, diese Verknüpfung wegzulassen. Stuellein und Kollegen (2016) stellen die Frage, ob diese vereinfachende Annahme für die Simulation von episodischen Gedächtnisaufgaben verworfen werden sollte (vgl. Kapitel 3).

Diese Frage kann beantwortet werden, indem wir die in Hofmann und Jacobs (2014) vorgeschlagenen Vorhersagen für EKP-Daten zu Rate ziehen. Der *bottom-up* getriebene Zugriff auf ein orthographisches Lexikon sollte demnach zwischen 100 und 200 ms beginnen (z. B. Sereno et al., 1998). Wenn eine experimentelle Manipulation vorliegt, die durch die semantische Ebene erklärt werden kann, dann lässt sich ein solcher Effekt am besten mit einer *Top-down*-Verknüpfung zwischen der semantischen und orthographischen Ebene des AROMs erklären.

Um festzustellen, ob die Annahme einer *Top-down*-Verknüpfung nötig ist, manipulieren wir in der EKP-Studie von Stuellein et al. (2016) die Anzahl assoziierter Wörter in einer episodischen Gedächtnisaufgabe (N = 29). In einer Studierphase werden 80 Wörter gelernt und sollen von den 80 neuen Wörtern in der Testphase unterschieden werden. Von den jeweils 80 studierten und nicht-studierten Wörtern haben jeweils 40 Wörter eine hohe Anzahl assoziierter Wörter im Reizmaterial, und die übrigen 40 Wörter haben jeweils eine geringe Anzahl assoziierter Wörter (Hofmann et al., 2011). Dann prüfen wir die aus Kapitel 2 abgeleitete Vorhersage für

neurokognitive Daten, dass ein EKP-Effekt im Zeitbereich zwischen 100 und 200 ms beginnt. Diese Hypothese lässt sich bestätigen. Bei studierten und nicht-studierten Wörtern mit einer hohen Anzahl assoziierter Wörter im Reizmaterial zeigt sich eine stärkere Positivierung der P200 ab 150 ms, was sich im Sinne einer Erleichterung des orthographisch-lexikalischen Zugriffs durch viele Assoziationen erklären lässt. Das heißt, zukünftige Versionen des AROMs sollten die vereinfachende Annahme dieser *Top-down*-Verknüpfung verwerfen (vgl. Kapitel 6.1). Darüber hinaus finden wir auch Effekte in der Komponente, die seit Kutas und Hillyard (1980) mit semantischen Repräsentationen verknüpft ist, der N400. Diesen Effekt würden wir mit semantischer Kompetition erklären, die durch die Anzahl assoziierter Wörter in einer episodischen Gedächtnisaufgabe verringert wird: Wenn ein Wort eine geringe Anzahl assoziierter Wörter im Reizmaterial hat, dann führt das zu verringerter (bzw. auch verkürzter) semantischer Kompetition (siehe auch Kapitel 6.1) und damit auch zu geringeren Negativierungen im EKP.

Darüber hinaus konnten wir die bereits bekannten Effekte der Anzahl von Assoziationen zu den anderen Stimuli in Fehler- und Ja-Antwortraten replizieren (Hofmann et al., 2011; Hofmann & Jacobs, 2014; Kuchinke et al., 2013) und um Reaktionszeiteffekte erweitern. Während eine hohe Anzahl assoziierter Wörter im Reizmaterial für nicht-studierte, neue Wörter zu langsameren Reaktionszeiten und höheren Fehlerraten führen, zeigen sich schnellere Reaktionen und höhere Erinnerungsleistungen bei alten Wörtern mit vielen Assoziierten. Dies lässt sich mit den in Grainger und Jacobs (1996) vorgeschlagenen, an die Theorie der Signalentdeckung angelehnten Entscheidungsmechanismen des MROMs für die lexikalische Entscheidungsaufgabe erklären (z. B. Broadbent, 1967; Jacobs, Graf, & Kinder, 2003): Basierend auf der Stärke der Aktivierung in einer frühen Stufe der Verarbeitung (*cycle* 1-7) werden Entscheidungskriterien gesetzt: Höhere Aktivierungen in den ersten Modellzyklen führen zu einem liberalen Entscheidungs-Kriterium für die Ja-Antwort, das heißt, ein relativ geringes Gedächtnissignal kann in den späteren Modellzyklen zu einer schnelleren positiven Antwort führen (vgl. Kapitel 6.1). Die Nein-Antwort wird in Jacobs und Grainger (1996) dann gegeben, wenn bis zu einer zeitlichen Deadline keine positive Antwort erreicht wurde. Wenn höhere Aktivierungen in diesen frühen Modellzyklen vorliegen, dann stellt dieser *Fast-guess*-Mechanismus fest (z. B. Braun et al., 2006), dass eine Ja-Antwort wahrscheinlich ist. Deshalb verzögert er die zeitliche Deadline. Der *Fast-guess*-Mechanismus gibt dem System mehr Zeit,

um doch noch eine positive Antwort zu erreichen (siehe aber Dufau, Grainger, & Ziegler, 2012; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008). Wenn wir den Versuchspersonen die Frage stellen, ob ein Wort studiert wurde, führt die hohe assoziative Aktivierung bei studierten Wörtern zu schnelleren und bei nicht-studierten Wörtern zu langsameren Antworten. Die von Jacobs und Grainger (1996) postulierten Entscheidungsmechanismen für die Frage, ob der dargebotene Stimulus ein Wort ist, lässt sich auch auf die Frage, ob der dargebotene Stimulus in der Studierphase gelernt wurde, übertragen.

4.4 Assoziatives und semantisches Priming in einer lexikalischen Entscheidungsaufgabe

McNamaras (2005) nahezu exhaustiver Überblick über semantisches Priming kommt zu dem Schluss, dass diese Literatur vor der Herausforderung steht „*auf irgendeine plausible Art und Weise zwei hochgradig assoziierte Wörter zu finden, die nicht semantisch relatiert sind*" (McNamara, 2005, p. 86[22]). Nach meiner Meinung ist die Schwierigkeit, solche theoretisch möglichen Wortpaare zu finden, der klassischen Definition der Assoziation zwischen zwei Wörtern geschuldet. Assoziatives Priming wird in klassischen experimentellen Paradigmen der Psychologie typischerweise durch menschliche Leistungen in der freien Assoziationsaufgabe definiert (z. B. Lucas, 2000; Roediger & McDermott, 1995), semantisches Priming hingegen beispielsweise durch Synonyme (Lucas, 2000). Diese operationalen Definitionen lassen sich im Prinzip auch durch das gemeinsame Auftreten von Wörtern in großen Korpora vorhersagen (Rapp, 2002; Wettler & Rapp, 1993).

Warum könnten operationale Definitionen, die auf anderen menschlichen Leistungen basieren, problematisch sein? Ein Problem liegt sicher darin, dass freie Assoziationen starke interindividuelle Unterschiede aufweisen (Jung-Merker & Rüf, 2011). Im Laufe der Leseentwicklung tendieren Kinder beispielsweise immer mehr dazu in dieser Aufgabe Wörter zu nennen, die auch eine semantische Relation aufweisen – vorher nennen Kinder hier insbesondere Wörter, die häufig gemeinsam mit dem gegebenen Wort auftreten (Nelson, 1977; vgl. Franke et al., 2017). Die multivariate operationale Definition der Assoziation durch die freie Assoziationsaufgabe zur Vorhersage von lexikalischen Entscheidungsaufgaben (vgl. z. B. Hutchison, 2003; Lucas, 2000) führt zu dem Problem, dass – je nach Stichprobe in der freien Assoziationsaufgabe – assoziative und semantische Variablen mehr oder weniger stark konfundiert sein können (vgl. z. B. Hofmann et al., 2018). Deshalb erscheint es schwer, Wörter zu finden, die von Erwachsenen in der freien Assoziationsaufgabe genannt werden und zu dem Zielwort in keiner semantischen Relation stehen (McNamara, 2005, p. 86). Sollten wir lieber Assoziationsnormen von Kindern in einem frühen Stadium der Leseentwicklung verwenden, die eher nach dem Prinzip der Kontiguität erzeugt werden (Nelson, 1977)? In jedem Fall müssten wir uns dann fragen, wovon unsere „unabhängigen Variablen" denn unabhängig sein

[22] „Having devoted a fair amount of time perusing free-association norms, I challenge anyone to find two highly associated words that are not semantically related in some plausible way."

sollen. Sie sind jedenfalls nicht unabhängig vom menschlichen Verhalten in der spezifischen Normstichprobe definiert.

Die in Kapitel 4.4 dargestellten Untersuchungen verfolgen zwei Ziele:

Zum einen prüfen wir, ob das in Abbildung 1A dargestellte Modell neben Assoziations-*ratings* und episodischen Gedächtnisaufgaben auch reproduzierbare Verhaltens-Vorhersagen für gebahnte lexikalische Entscheidungsaufgaben machen kann. Insbesondere testen wir, ob assoziative und semantische Priming-Effekte in der klassischen Literatur theoretisch sparsam durch ein einziges Prinzip erklärt werden können (Hutchison, 2003; Lucas, 2000): Assoziatives Priming, das in klassischen Studien durch die freie Assoziationsaufgabe operationalisiert wird, sollte sich durch die direkte AS zwischen Prime- und Zielwort abbilden lassen; semantisches Priming, das zum Beispiel durch die Verwendung von relativ synonymen Prime-Zielwort-Paaren operationalisiert wird (Lucas, 2000), sollte sich hingegen durch die Anzahl gemeinsamer Assoziierter abbilden lassen (Bordag, 2007). Dieses zweifaktorielle Vorgehen beinhaltet eine Zelle aus Wortpaaren mit einer hohen AS und wenigen gemeinsamen Assoziierten und erfüllt damit McNamaras (2005, p. 86) Herausforderung.

Zum anderen sollten die in Abbildung 1B dargestellten *model-to-data connections* für fMRT-Daten geprüft werden. Bei der ersten Einreichung dieser Studie wurden beide Ziele miteinander verknüpft. Die Gutachter schlugen auf Grund der unterschiedlichen Befundmuster der Verhaltens- und fMRT-Daten jedoch vor, die beiden Ziele in zwei getrennten Veröffentlichungen zu verfolgen – einen Vorschlag, den wir angenommen haben. Im Folgenden werden zuerst die bereits veröffentlichten Befunde angeführt. Dann werden die von Roelke, Franke, Radach, Jacobs und Hofmann (2016) eingereichten fMRT-Befunde kurz skizziert.

Publikation 6: Roelke, Franke et al. (2018)

Roelke, A., Franke, N., Biemann, C., Radach, R., Jacobs, A. M., & **Hofmann**, M. J. (2018). A novel co-occurrence-based approach to predict pure associative and semantic priming. *Psychonomic Bulletin and Review*, *25*(4), 1488–1493.

In dieser Untersuchung prüfen wir auf Basis von Verhaltensdaten die Hypothesen, dass die direkte Assoziation zwischen Bahnungs- und Zielwort assoziative Priming-Effekte erklärt; und dass viele gemeinsame Assoziierte semantischen Priming-Effekte in der lexikalischen Entscheidungsaufgabe erklären. Um diese Hypothesen zu untersuchen, manipulieren wir neben direkten und gemeinsamen Assoziationen den dritten experimentellen Faktor der SOA. Die Prime-Wörter werden 150 ms dargeboten, und es wird entweder für 50 ms oder für 850 ms ein weißer Bildschirm präsentiert, bevor die Zielwörter dargeboten werden; somit vergleichen wir eine SOA von 200 ms mit einer SOA von 1000 ms. In der klassischen Priming-Literatur zeigte sich nämlich, dass antworterleichternde assoziative Effekte bei beiden SOAs auftreten (Hutchison, 2003; Lucas, 2000; Plaut & Booth, 2000). Ein etwas weniger eindeutiges Befundmuster zeigt sich für semantisch relatierte Prime-Zielwort-Paare: Während antworterleichternde semantische Priming-Effekte hauptsächlich bei kurzer SOA (< 250 ms) auftreten (Ferrand & New, 2003; Lucas, 2000; McNamara, 2005), können bei langer SOA (> 500 ms) sowohl antworterleichternde als auch antworterschwerende Effekte auftreten, je nachdem ob eine Erwartung erfüllt wird oder ob die Antwort durch semantische Kompetition verzögert wird (Neely, 1977; Plaut & Booth, 2000). Deshalb erwarten wir Effekte direkter Assoziationen bei beiden SOAs, während wir die Effekte gemeinsamer Assoziierter eher für die kurze SOA erwarten. Die Verhaltensergebnisse in dem behavioralen und dem fMRT-Experiment stützen diese Hypothesen (je $N = 32$): Es zeigen sich schnellere Reaktionszeiten bei direkt assoziierten Wörtern in beiden SOAs; die Effekte gemeinsamer Assoziierter beobachten wir jedoch ausschließlich bei der kurzen SOA. Dieses Befundmuster stützt die Hypothese, dass sich assoziative Priming-Effekte durch die AS und semantische Priming-Effekte durch die gemeinsamen Assoziierten abbilden lassen. Auch letztere lassen sich also theoretisch sparsam durch eine 0/1-kodierte AS erklären.

Roelke et al. (2016)

Roelke, A., Franke, N., Radach, R., Jacobs, A. M., & **Hofmann**, M. J. (2016). *Semantic higher order but not direct associations prime ventral visual stream activations.* Manuskript zur Veröffentlichung eingereicht, 1-26.

Die damals eingereichten (Roelke et al., 2016) und sich derzeit wieder für den Veröffentlichungsprozess in Vorbereitung befindenden fMRT-Befunde zeigen ein anderes Befundmuster als Roelke, Franke et al. (2018). Während die Verhaltensdaten sowohl für direkte Assoziationen zwischen Prime- und Zielwort als auch für gemeinsame Assoziierte Effekte hervorbringen, finden wir keine fMRT-Effekte für direkte Assoziationen. Wie ließe sich dieses negative Ergebnis erklären?

Eine Möglichkeit wäre anzunehmen, dass die direkten Assoziationen zwischen jeweils zwei Repräsentationen im Sinne verkörperter semantischer Repräsentationen im Gehirn eine hohe räumliche Variabilität aufweisen (z. B. Patterson et al., 2007; Pulvermüller & Fadiga, 2010). Wenn ein Wort, das sich auf eine Handlung bezieht, im Motorkortex repräsentiert wird und ein Nomen, das sich auf ein Klang-relatiertes Wort bezieht, in superior temporalen Arealen (STL), dann müsste die Assoziation diese beiden Areale betreffen. Die assoziative Relation zwischen zwei auf auditive Inhalte bezogene Nomen bliebe jedoch innerhalb des Temporallappens (vgl. z. B. Trumpp, Traub, & Kiefer, 2013). Variabilität in der räumlichen Kodierung spezifischer Assoziationen könnte somit erklären, warum wir keine fMRT-Effekte für die AS beobachten können. Deshalb sollte der Effekt direkter Assoziationen zwischen Prime- und Zielwörtern nochmals innerhalb einer einzigen verkörperten Repräsentation untersucht werden: zum Beispiel direkte Assoziationen zwischen ausschließlich Klang-relatierten Begriffen; oder zwischen zwei konstant gehaltene verkörperte Repräsentationen im Gehirn, zum Beispiel Handlungs-relatierte Prime-Wörter und Klang-relatierte Zielwörter. Hierzu ließen sich auf Basis des AROMs Wörter auswählen, die eine hohe Anzahl gemeinsamer Assoziierter mit „Klang" oder „Handlung" aufweisen (Pulvermüller & Fadiga, 2010; Kiefer & Pulvermüller, 2012). Auf Basis dieses symbolisch repräsentierenden Modells ließen sich also Hypothesen verkörperter Repräsentationen leicht in eine komputational berechenbare, symbolisch repräsentierende Operationalisierung überführen.

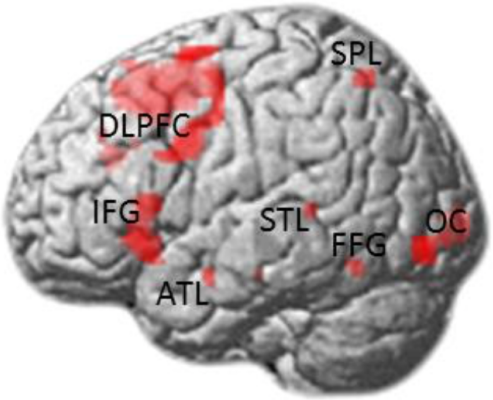

Abbildung 2. Signifikante Aktivierungserhöhungen aus Roelke et al. (2016) für Zielwörter mit wenigen im Vergleich zu vielen gemeinsamen Assoziierten mit dem Bahnungsreiz. Für weitere Erläuterungen siehe Fließtext.

Die Anzahl gemeinsamer Assoziierter von Prime- und Zielwort löst Effekte in allen Regionen aus, die in Abbildung 1B vorhergesagt wurden (Abb. 2). Wörter mit vielen gemeinsamen Assoziierten lösen geringere Aktivierungen im OC, FFG und STL sowie im linken und rechten inferioren Frontalgyrus aus (vgl. Tabelle 3, Roelke et al., 2016). Somit stützen diese Befunde die in den Kapiteln 3 und 4.3 beschriebenen Evidenzen für vollständige Interaktivität zwischen den Repräsentationsebenen im AROM (vgl. Abb. 1).

Während sich der IFG-Effekt am besten durch verringerte semantische Kompetition erklären lässt, zeigen die Befunde im OC und FFG, dass die zum Einsparen komputationaler Ressourcen gemachte Annahmen fehlender *Top-down*-Verknüpfungen zu den Ebenen visueller Eigenschaften und orthographischer Repräsentationen auch für lexikalische Entscheidungsaufgaben verworfen werde sollten, wenn genügend komputationale Ressourcen für eine solche Simulation zur Verfügung stehen (vgl. Kapitel 6.1). Unter den gefundenen Regionen ist insbesondere der Effekt im ATL erwähnenswert, da diese Region als semantische Konvergenzzone diskutiert wird (Patterson et al., 2007, p. 977), in der die interaktiven semantischen Aktivierungen aus allen Modalitäten zusammenfließen (vgl. amodale Repräsentationen in Kiefer & Pulvermüller, 2012; Patterson et al., 2007). Dort sollte also der Effekt gemeinsamer Assoziierter entstehen und sich auf die anderen Regionen ausbreiten (vgl. Abb. 1). Der Effekt im STL ließe sich auf auditive Repräsentationen zurückführen (z. B. Price, 2000), was sich durch eine Interaktivität des im AROM abgebildeten lexikalischen Pfades zur Graphem-Phonem-Route von *Dual-route*-Modellen erklären ließe (Braun et al., 2009;

Coltheart et al., 2001; Perry et al., 2007). Der Effekt im dorsolateralen präfrontalen Cortex (DLPFC) könnte dadurch entstehen, dass orthographische Aktivierungen einen Einfluss auf die Kriteriensetzung in der lexikalischen Entscheidungsaufgabe haben (vgl. z. B. Grainger & Jacobs, 1996; Perea & Carreiras, 2003; Kapitel 6.1). Der DLPFC wird mit Kontrollprozessen der Kriteriensetzung in Verbindung gebracht (z. B. Kuchinke et al., 2011).

5. Lehrmaterialien in deutschsprachigen Sammelwerken

5.1 Neurokognitive Modellierung

Publikation 7: Jacobs & Hofmann (2013)

Jacobs, A. M., & **Hofmann**, M. J. (2013). Neurokognitive Modellierung. In E. Schröger & S. Koelsch (Eds.), *Enzyklopädie der Psychologie. Affektive und kognitive Neurowissenschaft* (pp. 431–447). Göttingen: Hogrefe.

Dieses Buchkapitel bietet gerade für Studierende der Psychologie einen relativ leicht lesbaren Einstieg in die neurokognitive Modellierung. Wir beginnen in Abschnitt 1 mit den drei wichtigsten Fragen der neurokognitiven Modellierung: Wo im Gehirn wird der neurokognitive Prozess abgebildet – eine Frage, die sich mit der fMRT-Methode beantworten lässt. Wann finden welche Prozesse statt? Diese Frage lässt sich mit Methoden, wie dem Elektroenzephalogramm (EEG), beantworten. Doch die für die neurokognitive Modellierung wichtigste Frage ist die Frage nach dem „Wie": Wie funktioniert ein neurokognitiver Prozess genau, und welche Vorhersagen lassen sich für das „Wo" und „Wann" sowie für das „behaviorale Endprodukt" des kognitiven Prozesses aus einem Modell ableiten.

In Abschnitt 2 fragen wir nach dem Nutzen und den Grenzen von Modellen und Theorien, die einen kognitiven Prozess nur verbal oder mit Kästchen und Pfeilen beschreiben. Dann skizzieren wir einige Beispiele von Modellen, die den Entwicklungsstatus rein präquantitativer Theoriebildung bereits hinter sich gelassen haben (z. B. Anderson et al., 2004; vgl. Kapitel 2).

Schließlich diskutieren wir den Zusammenhang von Gehirn und Geist im Sinne des *Reverse-inference*-Problems in Abschnitt 3 (Poldrack, 2006). Zum Beispiel zeigen viele Studien, dass angstinduzierende Stimuli neuronale Aktivierungen in der Amygdala hervorgebracht haben (z. B. Davis, 1992). Mehr oder minder implizit könnten andere Studien aus einer Aktivierung der Amygdala schließen, dass bei einem gegebenen Stimulus Angst vorlag. Diese umgekehrte Inferenz wäre zwar logisch invalide („Bejahung der Konsequenz"), aber nützlich zum Generieren neuer Hypothesen, wenn man sich der logischen Fallstricke und des bedingten Wahrscheinlichkeits-Charakters logischer Inferenzen im Bereich neurokognitiver Forschung bewusst ist (Poldrack, 2006). Wir können gemäß Poldrack (2006) bei einer gegebenen Aufgabe

a nur bedingt auf das Vorliegen eines kognitiven Prozesses c schließen: $p(c|a)$. Von diesem kognitiven Prozess ausgehend schließen wir wiederum nur probabilistisch auf das Vorliegen neuronaler Aktivierungen n: $p(n|c)$.

In Abschnitt 4 veranschaulichen wir die Beziehungen zwischen Aufgaben, kognitiven Prozessen und Aktivierungen. Wenn wir zum Beispiel an David Poeppels (1996) Frage denken, wo im Gehirn phonologische Prozesse stattfinden [$p(n|a)$], dann ist zu bedenken, dass Modelle möglichst genau spezifizieren sollten, in welchen Aufgaben phonologische Prozesse eine wie starke Rolle spielen [$p(c|a)$]. Im Bereich der visuellen Worterkennung rekrutieren Aufgaben der Benennung, lexikalischen Entscheidung oder perzeptuellen Identifikation sowohl gemeinsame als auch unterschiedliche kognitive Prozesse. Diese funktionelle Überlappung sollte von Modellen möglichst explizit gemacht werden (z. B. Coltheart et al., 2001; Grainger & Jacobs, 1996; Perry et al., 2007).

Für den Bereich der neurokognitiven Modellierung erscheint die Maximierung von $p(n|c)$ am wichtigsten. Hier bieten kognitive Ontologien, wie zum Beispiel *BrainMap* (Fox & Lancaster, 2002), die Möglichkeit Schlagwörter für kognitive Prozesse, wie *phonological processing*, auszuwählen und bildgebende Studien zu diesen Schlagwörtern zu finden. Die Nützlichkeit dieser Datenbanken ist jedoch begrenzt, weil viele Begriffe, wie zum Beispiel „*semantic*" oder „*working memory*", keine adäquate Beschreibung des kognitiven Prozesses liefern und damit unterspezifiziert sind. Darüber hinaus ist zu berücksichtigen, dass gute kognitive Ontologien Vorhersagen über die Koaktivierung verschiedener neuroanatomischer Regionen und die Aktivierungsverteilung machen sollten (Price & Friston, 2005).

Schließlich erörtern wir vier Modellevaluationskriterien in Abschnitt 5 für den deutschsprachigen Leser (vgl. Kapitel 2; Jacobs & Grainger, 1994): (i) eine möglichst genaue Beschreibung der Daten, (ii) Verallgemeinerbarkeit, (iii) Einfachheit und Falsifizierbarkeit und (iv) eine möglichst genaue Erklärung. Zur Veranschaulichung letzteren Punktes erklären wir den Begriff der Ad-hoc-Annahme, die genau das Phänomen beschreibt, für das sie eingeführt wurde. Erst wenn das Modell darüberhinausgehende, neue Phänomene erklären kann, gewinnt es an Erklärungswert. Dieses Prinzip lässt sich auch durch den folgenden Term veranschaulichen: Erklärungswert = Anzahl erklärbarer Phänomene / Anzahl dazu nötiger Annahmen. Wenn dieser Term gleich eins ist, besteht das Modell nur aus Ad-hoc-Annahmen. Es benötigt genauso viele Annahmen wie es empirische Phänomene beschreibt. Ein solches Modell erklärt noch nichts und kann somit als Tautologie gesehen werden.

5.2 Der Prozess des Lesens in interaktiven Aktivierungsmodellen

Publikation 8: Radach & Hofmann (2016)

Radach, R., & **Hofmann**, M. J. (2016). Graphematische Verarbeitung beim Lesen von Wörtern. In U. Domahs & B. Primus (Eds.), *Laut, Gebärde, Buchstabe (Handbuch Sprachwissen, Band 2)* (pp. 455–473). Berlin: De Gruyter Mouton.

Dieses Buch richtet sich allgemein an den wissenschaftlich interessierten Leser, der Laut-, Schrift- und Gebärdensprache aus dem Blickwinkel verschiedener Disziplinen betrachten möchte. Unser Buchkapitel führt dabei in Abschnitt 1 in die beiden Traditionslinien der Worterkennungs- und Leseforschung ein. Zum einen veröffentlichte James McKeen Cattell (1885) seine Abhandlung „*Über die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben*", die ihn zu einem der Mitbegründer der tachistoskopischen Worterkennungsforschung machte. Im Forschungsfeld der visuellen Worterkennung werden den Versuchspersonen einzelne Buchstabenfolgen dargeboten, und sie sollen sie zum Beispiel benennen oder entscheiden, ob es ein Wort oder Nichtwort ist. Zum anderen erschienen die ersten Arbeiten zum dynamischen Lesen nur wenige Jahre später. Beide Forschungslinien wurden bald zu einem Lehrbuch zusammengefasst (Huey, 1908).

Abschnitt 2 behandelt das visuelle Erkennen einzelner Wörter und deren neurokognitive Modellierung. Es werden die grundlegenden Ebenen des IAMs und AROMs sowie deren neuronale Entsprechungen eingeführt (Hofmann & Jacobs, 2014, vgl. Abb. 1, Kapitel 2). Schließlich findet sich dort auch Tabelle 1, in der die wichtigsten neurokognitiven Effekte der Worterkennung hinsichtlich der „Wo" und „Wann" Fragen eingehend konkretisiert werden. Ähnlich wie in Hofmann und Jacobs (2014) diskutieren wir Verhaltens-, EEG- und fMRT-Effekte der Länge, Sequenz, Frequenz, Wiederholung, orthographischen Nachbarschaft und „Semantik" (vgl. Kapitel 2). Schließlich zeigen wir auch eine erste exemplarische Simulation semantischer Aktivierungsausbreitung im Satzkontext mit dem AROM (Abb. 2, Radach & Hofmann, 2016). In dem Satz „Claudia hatte ihr Fahrrad auf der Straße stehen lassen" sind die Wörter „Fahrrad", „auf" und „der" mit dem Wort „stehen" assoziiert. Deshalb ist dieses Wort

bereits aktiviert, bevor es präsentiert wurde, was die OC-Aktivierungsunterschiede in Kapitel 3 erklärt.

In Abschnitt 3 beschäftigen wir uns mit der Dynamik des natürlichen Lesens in der Blickbewegungsforschung. Hierzu werden Begriffe, wie Sakkade, Fixationsdauer und Blickspanne, eingeführt, um die im Folgenden präsentierten Befunde zum dynamischen Lesen verstehen zu können. Es werden Methoden, wie die *boundary technique,* beschrieben, mit der man die Nutzung parafovealer Informationen beim Lesen untersuchen kann, indem man die Wörter in der Parafovea modifiziert oder maskiert. Der Bildschirm ändert sich, wenn eine unsichtbare Grenze vom Auge überschritten wird, bevor das Wort direkt foveal fixiert wird. So kann man den parafovealen *preview* getrennt von foveal verfügbaren Informationen untersuchen. Dann werden einige Experimente beschrieben, die zeigen welche parafovealen Informationen der Leser nutzt. Interessanterweise scheint der Leser keine extrafovealen Wortlängeninformationen zu nutzen (Inhoff, Radach, Eiter, & Juhasz, 2003). Dennoch hat die orthographische Regularität von Wörtern in der Parafovea einen Einfluss auf die Programmierung von Sakkaden (Radach, Inhoff, & Heller, 2004) – oder Kontexteigenschaften, wie die Vorhersagbarkeit des nächsten Wortes (Balota, Pollatsek, & Rayner, 1985).

Schließlich wird auch ein IAM des dynamischen Lesens vorgestellt, das Glenmore-Modell (Reilly & Radach, 2006). Im Gegensatz zur Ebene visueller Eigenschaften (McClelland & Rumelhart, 1981; Hofmann et al., 2011), wird der visuelle Input hier relativ zur aktuellen Blickposition bestimmt: Der Input besteht aus einem 30 Buchstaben- und Leerzeichen umfassenden Vektor mit einem abfallenden Gradienten an übermittelter visueller Information. Die übertragene Information nimmt mit dem Abstand zur aktuellen Blickposition ab (Abb. 3, Radach & Hofmann, 2016). Wie im IAM wird diese Information an eine Ebene der Buchstaben weitergeleitet, die wiederum eine Ebene an orthographischen Wortrepräsentationen aktiviert, welche mit einer steigenden Wortfrequenz schneller an Aktivierung gewinnen können. Doch im Gegensatz zum IAM wirkt der visuelle Input auch auf eine Ebene an *saliency units*, die gleichzeitig auch Informationen von der Buchstabenebene erhalten. Auf der Ebene orthographischer Wortformen inhibiert das erfolgreiche Wiedererkennen die im Wort enthaltenen Buchstabenrepräsentationen. Deshalb wird die visuelle Salienz der Buchstaben verringert, wenn das jeweilige Wort hohe Aktivierungen aufweist und damit erkannt wurde. So lässt sich die Interaktion basaler visueller Verarbeitungsprozesse mit der orthographischen Identifikation in diesen *saliency units* abbilden. Im Verlauf der Modell-Dynamik erhalten

bestimmte Positionen im Blickfeld so jeweils mehr oder weniger Salienz. Darüber hinaus aktivieren die Buchstabeneinheiten das sogenannte Fixationszentrum. Dort werden die Aktivierungen der Buchstabeneinheiten aufsummiert. Somit überwacht das Fixationszentrum, ob noch hinreichend viele visuelle Informationen in den Buchstabeneinheiten analysiert werden müssen. Sobald diese summierte Aktivierung einen kritischen Wert unterschreitet, wird eine Blickbewegung ausgelöst und auf den Ort im Blickfeld programmiert, der in diesem Moment die höchste Salienz aufweist. Hier wird ein Ort gewählt, bei dem bereits hinreichend viele visuelle Input-Informationen vorliegen, um interessant zu sein. Auf Grund des abfallenden Gradienten relativ zur Blickposition kämen hierfür vor allem Positionen nahe zur aktuellen Blickposition in Frage. Für diese zentralen Positionen könnte die lexikalische Verarbeitung zu diesem Zeitpunkt jedoch bereits relativ abgeschlossen sein, insbesondere wenn es sich bei dem aktuellen Wort beispielsweise um ein häufiges Wort handelt. Deshalb werden zu diesem Zeitpunkt benachbarte Wörter mit hoher Wahrscheinlichkeit als das Ziel der Sakkade ausgewählt. Ist das aktuelle Wort hingegen ein seltenes Wort, wird die lexikalische Verarbeitung mit höherer Wahrscheinlichkeit noch nicht abgeschlossen sein, und für die Sakkade würde mit einer höheren Wahrscheinlichkeit ein Ziel im aktuellen Wort ausgewählt, insbesondere wenn es sich hierbei um ein längeres Wort handelt (vgl. Reilly & Radach, 2006).

6. Weiterführende Studien

6.1 Assoziative Aktivierungsausbreitung über Wortsequenzen

In den ersten Studien mit dem AROM haben wir die Assoziationen zwischen den Wörtern einer episodischen Gedächtnisaufgabe simuliert (Hofmann et al., 2011). Bei 160 Wörtern handelte es sich hier um eine relativ überschaubare Menge an Assoziationen. Dennoch stellt sich die Frage, welche experimentellen Konsequenzen aus einem AROM folgen müssten, wenn dieses mit einem vollständigen Lexikon implementiert würde. Ein solches AROM müsste theoretisch in der Lage sein, die Effekte morphologischer Familiengrößen abzubilden: Stamm-Morpheme, die mit einer größeren Anzahl Flexionen auftreten, werden leichter erkannt (z. B. Baayen, Dijkstra, & Schreuder, 1997). Da Rapp (2002) die Anzahl gemeinsamer Assoziierter dazu verwendet, Synonyme zu finden, sollten somit auch Priming-Effekte bei Synonymen erklärbar werden (vgl. Kapitel 4.1). Schließlich sollte sich mit einem vollständigen implementierten semantischen Lexikon auch die temporäre Aktivierung eines Bedeutungsfeldes sowie Effekte der Bedeutungsdominanz bei polysemen Wörtern simulieren lassen (z. B. Panchenko, Ruppert, Faralli, Ponzetto, & Biemann, 2017; Rodd, Gaskell, & Marslen-Wilson, 2002; Schvaneveldt, Meyer & Becker, 1976).

Nach dem ersten experimentellen Nachweis, dass sich gemeinsame Assoziierte dazu nutzen lassen, die Effekte semantischer Überlappung in gebahnten lexikalischen Entscheidungen quantitativ abzubilden (Kapitel 4.4; Roelke, Franke et al., 2018), erscheint ein AROM mit realistischeren Wortschatzgrößen als nächster logischer Schritt. Da alle Wörter und alle gemeinsamen Assoziierten von Roelke, Franke et al. (2018) mehr als die uns zur Verfügung stehenden etwa 50 GB Arbeitsspeicher und 100 GB Swap-Speicher benötigen, generieren wir für jedes Prime-Zielwort-Paar jeweils ein Lexikon mit Prime, Zielwort und allen Wörtern, die mit diesen assoziiert sind. Die resultierenden 200 Lexika variieren dabei zwischen 397 bis 23.470 Wortformen (M = 6.674, SD = 5.003). Um eine möglichst gute Vorhersage zu ermöglichen, erproben wir diese Simulation mit etwa 16.000 verschiedenen Parametersätzen. Dabei variieren wir die Skalierung der Exzitation, das heißt die Alpha-Parameter von der orthographischen zur semantischen Ebene und zurück, Alpha innerhalb der semantischen Ebene, den inhibitorischen Parameter Gamma innerhalb der semantischen Ebene sowie ein Gamma von der semantischen zur orthographischen Ebene (vgl. Abb. 1A). Darüber hinaus

werden verschiedene *Decay*-Parameter, das heißt Gedächtniszerfallsraten auf der semantischen Ebene, in Betracht gezogen (vgl. Hofmann et al., 2011).

Für die Selektion des im Folgenden präsentierten Parametersatzes werden als *model-to-data connections* drei verschiedene Prädiktoren aus dem Modell abgeleitet: Die *global lexical activation* (GLA) entspricht der summierten Aktivierung aller aktivierten, das heißt die Aktivierungsschwelle von 0 überschreitenden, orthographischen Einheiten über die ersten sieben Modellzyklen (Grainger & Jacobs, 1996). Diese wurde beispielsweise für die Vorhersage der N400 bei Nichtwörtern verwendet (Braun et al., 2006). Ein zweites Maß ist die *associative memory signal strength* (AMSS), wie wir sie für die Simulation episodischer Gedächtniseffekte eingeführt hatten (Hofmann et al., 2011). Sie entspricht der mittleren Aktivierung der semantischen Einheit des präsentierten Wortes innerhalb der ersten sieben Modellzyklen. Schließlich nehmen wir an, dass lexikalische Entscheidungen auch bei vorgeschaltetem Bahnungswort immer noch im Wesentlichen auf Grund der Aktivierungen auf der orthographischen Ebene gefällt werden. Für die Simulation von Reaktionszeiten werden deshalb die im MROM vorgeschlagenen Entscheidungsmechanismen übernommen (Grainger & Jacobs, 1996): Zum einen kann das Modell ein Wort identifizieren, wenn eine orthographische Einheit die Identifikationsschwelle (hier 0,6) überschreitet. Zum anderen wirkt der sogenannte *Fast-guess*-Mechanismus auf die Kriteriensetzung (z. B. Braun et al., 2006): Auf Basis früher orthographischer Aktivierungen wird ein Entscheidungskriterium für die späteren summierten orthographischen Aktivierungen gesetzt. Wenn diese frühe summierte orthographische *Fast-guess*-Aktivierung (hier über die ersten vier Modellzyklen) einen Wert von 0,29 überschreitet, dann wird das Kriterium für die summierte Aktivierung orthographischer Einheiten auf 0,72 anstatt auf 5,9 gesetzt. Eine früh klarwerdende, orthographische Vertrautheit führt zum Setzen eines liberalen Entscheidungskriteriums für das Summen-Kriterium der Ja-Antwort. Wird diese Summen-Aktivierung in einem Zyklus überschritten, klassifiziert das Modell den Reiz als hinreichend vertraut, um damit eine Ja-Antwort auszulösen. Auf Basis dieser Annahmen ergibt sich aus dem Modell eine simulierte Reaktionszeit. Der im Folgenden präsentierte, exemplarische Parametersatz weist das maximale Produkt aus simulierter Reaktionszeit, GLA und AMSS auf. Als abhängige Variable verwenden wir die aus Kapitel 4.4 vorliegenden *Item-level*-Reaktionszeiten ($N = 64$).

Der ausgewählte Parametersatz weist ein Alpha von der orthographischen zur semantischen Ebene von 0,09 auf. Der exzitatorische Alpha-Parameter von der semantischen zur orthographischen Ebene beträgt 0,12. Exzitation und Inhibition innerhalb der semantischen Ebene sind jeweils mit 0,15 und 0,09 gewichtet. Die Inhibition von der semantischen zur orthographischen Ebene bleibt 0, und der *Decay*-Parameter bleibt 0,07, wie in der orthographischen Ebene des originalen IAMs (McClelland & Rumelhart, 1981). Bei diesem Parametersatz ergibt sich für die GLA eine Korrelation von r = -0.23 mit den Reaktionszeiten. Für die AMSS ergibt sich eine Korrelation von r = -0.35. Die simulierten und die empirischen Reaktionszeiten weisen eine Korrelation von r = 0.39 auf. Verwendet man alle Prädiktoren in einer multiplen Regression, kommen wir auf eine Varianzaufklärung von r^2 = 0.20. Wenn wir die Items der acht experimentellen Bedingungen aus Roelke, Franke et al. (2018) jeweils in einem Mittelwert zusammenfassen und mit den simulierten Ergebnissen vergleichen, dann ergibt sich eine gute Passung von Modell und Daten (Abb. 3).

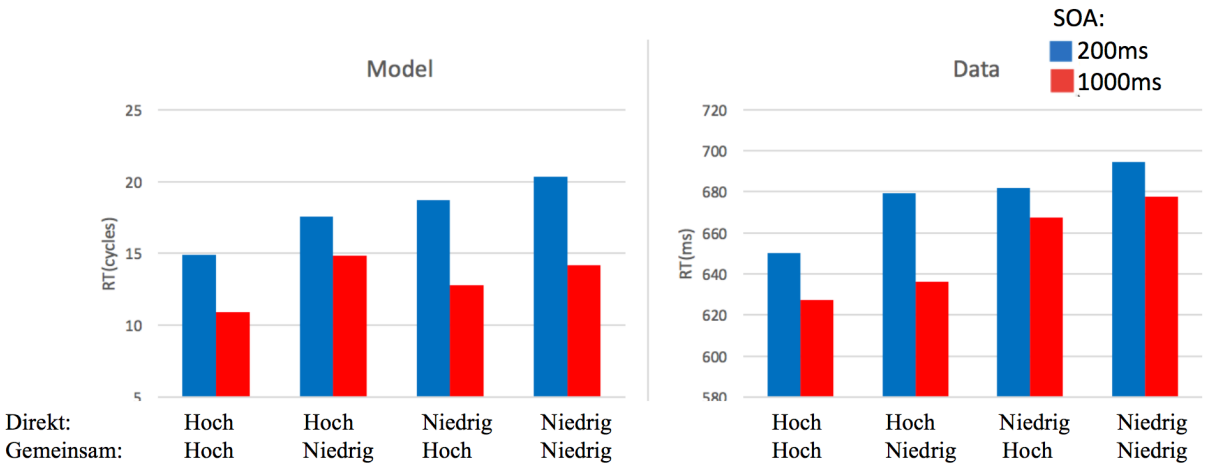

Abbildung 3. Simulierte (links) und empirische Reaktionszeiten (rechts) für die Stimuli aus Roelke, Franke et al. (2018). Die simulierten Reaktionszeiten entsprechen den Abbruchzyklen der lexikalischen Suche des Modells nach Erreichen der kritischen Aktivierungen (*cycles*). Das Modell kann einen antworterleichternden Effekt langer SOA (rot) im Vergleich zu kurzer erklären (blau). Darüber hinaus erklärt es antworterleichternde Effekte von direkten und gemeinsamen Assoziationen (vgl. Fließtext).

Bemerkenswert erscheint insbesondere, dass ein dynamisches AROM eine sehr gute Erklärung für den antwortbeschleunigenden Effekt einer langen SOA bietet: Wie in den empirischen Daten, führt eine lange SOA zu deutlich schnelleren Reaktionszeiten des Modells. Dies lässt sich dadurch erklären, dass die Prime-Aktivierungen bei mehr zur Verfügung stehender Zeit für assoziative Aktivierungsausbreitung zu generell höheren Aktivierungen führen, und somit

die Antwortkriterien schneller erreicht werden. Ebenso zeigt sich, dass direkte und/oder gemeinsame Assoziationen zwischen Prime- und Zielwort bei einer kurzen SOA die Antwort erleichtern, wenn man diese jeweils mit der Bedingung unverknüpfter Wörter vergleicht. Diese Effekte finden sich auch in den empirischen Daten (vgl. Roelke, Franke et al., 2018; Kapitel 4.3). Für die lange SOA beobachten wir innerhalb der nicht direkt assoziierten Prime-Zielwort-Paare, dass für die Wörter mit vielen gemeinsamen Assoziierten eine Antwortbeschleunigung im Vergleich zu den Wörtern mit wenigen gemeinsamen Assoziierten auftritt (Roelke, Franke et al., 2018). Innerhalb der direkt assoziierten Wörter zeigt sich bei langer SOA im Modell ein wesentlich deutlicherer Unterschied zwischen Wörtern mit hoher und geringer Anzahl gemeinsamer Assoziierter als in den in Abbildung 3 gezeigten Daten. Im vorliegenden Parametersatz weicht das simulierte Ergebnismuster nur in einer der acht Bedingungen stark von den empirischen Daten ab: Ausschließlich direkt assoziierte Prime-Zielwort Paare lösen bei langer SOA im Modell relativ langsame Reaktionszeiten aus. Dies führt innerhalb der Wörter mit direkter Assoziation zu einem wesentlich stärkeren Unterschied zwischen Wörtern mit vielen und wenigen gemeinsamen Assoziierten. Das Auftreten semantischer Priming-Effekte bei langer SOA erscheint als eine theoretische Möglichkeit, die sich manchmal auch empirisch nachweisen lässt (vgl. z. B. Rossel, Price, & Nobre, 2003; aber vgl. Plaut & Booth, 2000).

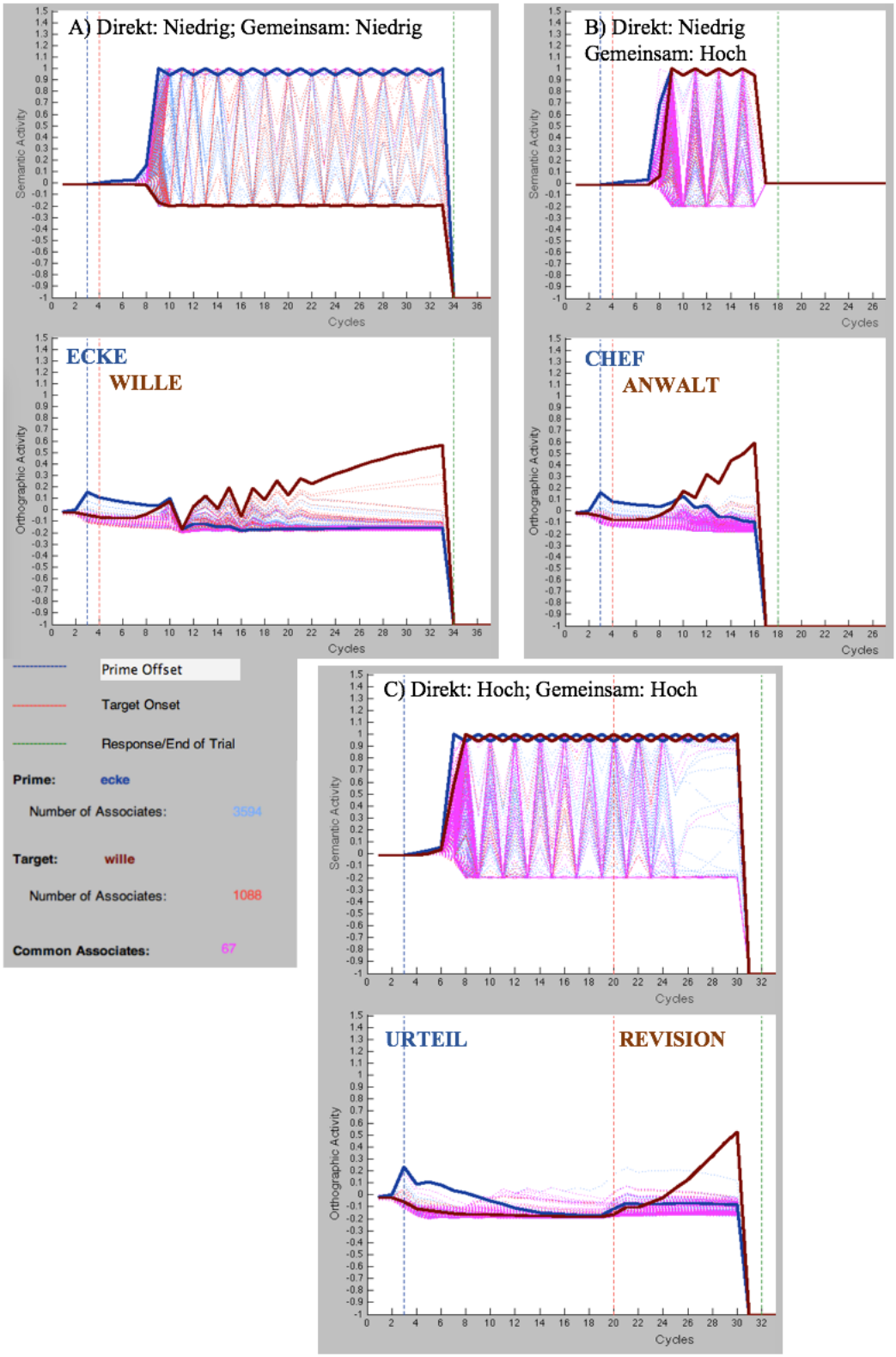

Abbildung 4. Drei Beispielsimulationen der dynamischen Aktivierungsausbreitung über Modellzyklen (X-Achse) der orthographischen (untere Abbildungen) und semantischen Worteinheiten auf den Y-Achsen (obere Abbildungen). In der oberen Reihe finden sich zwei Beispiele mit kurzer SOA (**Panel A** und **B**). In der unteren Abbildung ist eine Simulation mit einer langen SOA dargestellt (**Panel C**). Für weitere Erläuterungen siehe Fließtext.

In **Abbildung 4A** wird beim Simulationszyklus 1 auf der X-Achse beispielsweise der Prime „ECKE" dargeboten (dunkelblaue, fettgedruckte Kurve). Da ein Modellzyklus jeweils 50 ms der experimentellen Prozedur abbildet, endet der Prime nach Zyklus 3 (blaue vertikale, gestrichelte Linie), während dem Modell ab Zyklus 4 (SOA = 200 ms) das Zielwort „WILLE" (weinrote fettgedruckte Kurve) dargeboten wird, das weder eine direkte Assoziation noch viele gemeinsame Assoziierte mit dem Bahnungsreiz aufweist. Auf der semantischen Ebene entsteht etwa ab Zyklus 8, wenn die *Bottom-up*-Aktivierung des Zielwortes die semantische Ebene erreicht, eine starke Kompetition, die sich mit Aktivierungen im IFG in Zusammenhang bringen lässt (Kapitel 2; Hofmann & Jacobs, 2014, Abschnitt 4.5; Kapitel 4.4; Roelke et al., 2016; vgl. Thompson-Schill & Botvinick, 2006). Zuerst werden die Aktivierungsfunktionen aus der Kohorte assoziierter Wörter des Prime-Wortes (gepunktet hellblau) aktiv (vgl. Rahman & Melinger, 2009). Weil dieses allerdings nicht mehr präsentiert wird, erhöht sich etwa ab Zyklus 15 die Aktivierung der assoziierten Wörter des Zielwortes (gepunktet hellrot). Auf Grund der deutlich geringeren Menge an Assoziierten des Zielwortes (1088) im Vergleich zu den Assoziierten des Prime-Wortes (3594), verlieren erstere die Kompetition jedoch, und das Zielwort erreicht auf der semantischen Ebene zu keinem Zeitpunkt einen aktiven Zustand. Diese lang andauernde Kompetition überträgt sich auf die orthographische Ebene, was sich an den oszillierenden Aktivierungen etwa zwischen Zyklus 10 und 22 zeigen lässt. Schließlich kann sich am Ende auf Grund der anhaltenden *Bottom-up*-Aktivierung dennoch die orthographische Aktivierungsfunktion des präsentierten Zielwortes durchsetzen und erreicht die zur Identifikation notwendige Schwelle im Zyklus 33 nach Prime-Präsentation.

Abbildung 4B zeigt eine Beispielsimulation, bei der zwischen Prime- und Zielwort keine direkte Assoziation vorliegt. Stattdessen weisen die beiden Wörter viele gemeinsame Assoziierte auf (Aktivierungsfunktionen in gepunktetem Magenta). Diese semantische Überlappung, die in diesem Fall aus 1601 gemeinsamen Assoziierten besteht, führt nach Zielwortpräsentation zu einer relativ kurz andauernden semantischen Kompetition, wobei sich bei so kurzer Prime-Zielwort-Folge insbesondere die gemeinsamen Assoziierten relativ schnell durchsetzen. Die semantischen Eigenschaften von Prime- und Zielwort lassen sich relativ schnell semantisch integrieren (z. B. Dambacher et al., 2006): Bereits nach wenigen Oszillationen erhalten viele gemeinsame Assoziierte einen aktiven Zustand. In diesem aktiven Zustand wirkt kaum mehr semantische Inhibition von vermeintlichen stärker aktivierten Wörtern: Es scheint eine sehr hohe Anzahl aktiver, gleichermaßen erwartbarer Worteinheiten

zu geben. Insbesondere die Anzahl hochgradig aktiver gemeinsamer Assoziierter scheint theoretisch unlimitiert, solange die Anzahl an aktivierungsunterstützenden gemeinsamen Assoziierten nur hoch genug ist. Dies bildet eine symbolische Variante eines in sich relativ geschlossenen *cell assembly's* ab (vgl. Kapitel 1.1, Hebb, 1949, p. 60). In algorithmisch konkreten Simulationsmodellen lässt sich die kognitive Prozess-Dynamik eines solchen, relativ geschlossenen Systems aus symbolischen Einheiten meines Wissens ausschließlich mit dem AROM simulieren, da sich in diesem nonlinearen System alle erwartbaren Wörter nicht auf eine Gesamtwahrscheinlichkeit von 1 aufaddieren müssen (vgl. z. B. Mikolov, Chen et al., 2013; Rabovsky, Hansen, & McClelland, 2018). Es scheint kaum mehr Inhibition effektiv wirken zu können, wenn ein semantisches Feld einmal hinreichend aktiv ist. Wahrscheinlich deshalb finden beispielsweise Luke und Christianson (2016) in ihrer Blickbewegungsstudie keine inhibitorischen Effekte semantisch stärker aktivierter Nachbarn. Schließlich führt das auf der semantischen Ebene hochgradig erwartete Zielwort auf der orthographischen Ebene – nach kurzer oszillierender Kompetition – zu einer starken Exzitation des Zielwortes, was sich in einem relativ steilen Anstieg der orthographischen Aktivierungsfunktion des präsentierten Reizes äußert und damit auch zu einem relativ schnellen Erreichen der Identifikationsschwelle führt. Diese schnelle Modell-Reaktionszeit sollte im Vergleich mit der in Abbildung 4A gezeigten Simulation den Effekt der gemeinsamen Assoziierten erklären.

In **Abbildung 4C** findet sich ein Beispiel mit langer SOA. Um die Reaktionszeiten mit der kurzen SOA besser vergleichen zu können, wurde der Zeitpunkt der Zielwortpräsentation an dieselbe X-Achsenposition wie für Abbildung 4B geschoben. Prime- und Zielwort weisen zum einen eine direkte Assoziation auf, haben aber zum anderen auch 842 gemeinsame assoziierte Wörter. Deshalb erreicht das Zielwort bereits vor seiner Präsentation einen aktiven Zustand nahe 1 auf der semantischen Ebene. Bei langer SOA beginnt die semantische Kompetition bereits vor der Präsentation des Zielwortes nachzulassen: Eine Reihe erwartbarer Wörter nähert sich asymptotisch der Aktivierung von 1. Gleichzeitig sind bei Zielwortpräsentation hinreichend viele Assoziierte auf der semantischen Ebene aktiv, um den allgemein antworterleichternden Effekt langer SOA zu erklären: Das assoziative Netzwerk befindet sich nach langer SOA in einem energetisch angereicherten Zustand. Die verringerte semantische Kompetition lässt sich auch an der nicht oszillierenden orthographischen Aktivierungsfunktion nach der Zielwortpräsentation beobachten. So kann ein assoziativ mit dem Prime-Wort stark verdrahtetes Zielwort noch schneller wiedererkannt werden, wenn eine relativ lange Zeit zwischen diesen Reizen vergeht (vgl. Abb. 3B vs. 3C).

Während Collins und Loftus (1975) die assoziative Aktivierungsausbreitung als präquantitatives Modell in einem Taxonomie-basierten Netzwerk vorgeschlagen haben, verwenden wir mit der AS ein recht einfach berechenbares Maß für frequenzgewichtete Kontiguität, um damit die Dynamik der assoziativen Aktivierungsausbreitung im semantischen LZG besser zu verstehen. Semantische Ähnlichkeit ergibt sich in diesem Modell über gemeinsame Assoziierte, die sich in ihrer Aktivierungsdynamik mit dem AROM beschreiben lassen. Solche Simulationen werden erst in den letzten Jahren möglich, da erst jetzt hinreichend leistungsfähige Computer allgemein verfügbar sind. Es hat uns ein ganzes Jahr gekostet, die Matlab-Architektur des originalen AROMs an so große Lexika anzupassen und die Lexika des ersten Experimentes aus einer SQL-Datenbank auszulesen. Ein weiteres Rechenjahr wurde auf die Simulation des Experimentes mit etwa 16.000 Parametersätzen verwendet. Schließlich haben wir ein weiteres halbes Rechenjahr darauf verwendet, die Lexika für alle 6.644 Prime-Zielwort-Paare des *English lexicon projects* auszulesen (Hutchison et al., 2013), so dass wir bald sehen können, in wie weit sich unsere Simulationen verallgemeinern lassen und welche weiteren Effekte wir mit dem Modell abbilden können. Während solche vollständig symbolisch repräsentierenden Modelle sehr gut geeignet sind, um die neurokognitive Prozessdynamik abzubilden, sind verteilt repräsentierende *Hidden-unit*-Modelle, wie zum Beispiel das Modell von Elman (1990) und seine Nachfolgemodelle im Bereich des maschinellen Lernens (Mikolov, 2012; Mikolov, Chen et al., 2013), immer noch State of the Art, wenn es um die Simulation von „abstrakten" Lernprozessen geht (vgl. z. B. Grainger & Jacobs, 1998, p. 32). Zwar haben wir in Hofmann et al. (2011) den Beitrag von semantischen LZG-Assoziationen auf episodische Gedächtnisaufgaben simuliert. Dennoch kann dieses Modell keine neuen Assoziationen aus dem episodischen Kontext „mit"-lernen, denn das AROM besteht aus einem unveränderlichen LZG, in dem automatische Aktivierungs-Ausbreitungs-Prozesse simuliert werden. Um neue Informationen lernen zu können, bedarf es einer zusätzlichen episodischen Ebene, welche die semantischen Eigenschaften verschiedener Wörter vergleichen, ergänzen und trennen kann. Diese episodische Ebene würde neue symbolische Einheiten als assoziative Konjunktion semantischer Eigenschaften abbilden und damit – neben dem AROM – den zweiten Teil der *complementary learning systems* (CLS) darstellen (Kumaran & McClelland, 2012; Marr, 1970, 1971). Die so zu implementierende Kontrolle des automatischen Aktivierungsausbreitungsprozesses sollte insbesondere längerfristige Priming-Prozesse bei langer SOA besser erklären können (vgl. z. B. Mecklinger, 2010; Schneider & Shiffrin, 1977).

6.2 Das episodische Gedächtnis als komplementäres Lernsystem

Publikation 9: Hofmann & Kuchinke (2015)

Hofmann, M. J., & Kuchinke, L. (2015). "Anything is good that stimulates thought" in the hippocampus: Comment on "The quartet theory of human emotions: An integrative and neurofunctional model" by S. Koelsch et al. *Physics of Life Reviews*, *13*, 58–60.

Die Quartett-Theorie menschlicher Emotionen identifizierte vier verschiedene neuronale Emotionssysteme (Koelsch et al., 2015). Während wahrscheinlich viele Forscher mit ihr übereinstimmen würden, dass spezifische Kerngebiete im Hirnstamm, das Diencephalon und der OFC im Wesentlichen eine affektive Funktion innehaben, ist die These wohl schwerer zu stützen, dass der kleinste gemeinsame Nenner dessen, was der Hippocampus tut, eine affektive Funktion ist. Man könnte wahrscheinlich eher behaupten, dass er „Gedanken stimulieren" würde (Hyde, 1962, p. 108), die zu Inferenzen führen können (z. B. Kumaran & McClelland, 2012). Diese Erkenntnisfunktion entspricht wahrscheinlich genau dem Typus (positiver) Emotion, für die hippokampale Prozesse hinreichend sein könnten. Deshalb haben wir in den Titel dieses Kommentares das Zitat von Oscar Wilde integriert (Hyde, 1962).

Was sind die wesentlichen Funktionen des Hippocampus? Zum einen fließen LZG-Informationen aus dem cerebralen Cortex in ihn ein (Rolls, 2007). Da das AROM ein Modell des LZG ist, lässt es sich als eine Implementierung dieses Teils der CLS-Theorie begreifen (Marr, 1970, 1971; McClelland, McNaughton, & O'Reilly, 1995). Wir könnten zum Beispiel annehmen, dass viele direkte Assoziationen zwischen den Wörtern des Idioms „*weapons of mass destruction*" vorliegen, die als LZG-Information in den Hippocampus fließen. Dabei liegt die erste Funktion dieser syntagmatischen Relationen darin, teilweise aktivierte Muster zu vervollständigen (Yassa & Stark, 2011). Hätten wir einen Teil dieser Phrase nicht verstanden, würde wahrscheinlich über das Subikulum Aktivierung in den Cortex zurückgesendet, um das unvollständige Muster zu ergänzen (Rolls, 2007). Gelingt auch dann kein erfolgreicher Abgleich mit dem LZG, dann wirkt die zweite, zur Musterergänzung komplementäre Funktion: die Mustertrennung. Seit Eriksson und Kollegen (1998) gezeigt haben, dass auch im *dentate gyrus* (DG) des erwachsenen Hippocampus neue Zellen erzeugt werden, ist eine Debatte darüber entbrannt, ob und wie viele neue Neuronen dort bei Erwachsenen von ihren

Ahnenzellen „geboren" werden (Snyder, 2019; Sorrells et al., 2018). Kumaran und McClelland (2012) postulieren, dass dort die sogenannten *conjunction units* eine neue Verknüpfung bestehender, symbolischer Wissenseinheiten abbilden. Wie entstehen möglichst solide, neue Gedächtnis-Engramme? Mit der CLS-Theorie ließe sich vorhersagen, dass eine optimale Kombination aus bestehenden LZG-Informationen und neuen „Konjunktionen" eine optimale Gedächtnisenkodierung begünstigt. Als Beispiel könnte man hier die Wortsequenz „*weapons of mass distraction*" nennen (Giora, Kronrod, Elnatan, Shuval, & Zur, 2004), die sich aber vom oben genannten Idiom in einem wesentlichen Merkmal unterscheidet und damit auch den Prozess der Mustertrennung stimuliert.

Hier schließt sich auch der Kreis zu Theorien der Emotion: „*Weapons of mass distraction*" ist Teil des Titels eines Artikels, der eine Synthese aus Ästhetik-Theorien bildet, die behaupten, dass entweder Vertrautes oder Neues besonders schön sei: Giora und Kollegen (2004) schlugen vor, dass etwas dann besonders schön ist, wenn es optimal innovativ ist. In der ersten Förderphase eines Drittmittelprojektes[23] haben wir das in Kapitel 6.1 präsentierte Modell des semantischen LZG weiterentwickelt. In der derzeit laufenden Verlängerung haben wir es uns zum Ziel gesetzt dieses Modell um eine episodische Ebene zu erweitern, welches aus *conjuction units* besteht (Kumaran & McClelland, 2012). Hier werden wir unter anderem die Hypothese prüfen, dass eine Kombination aus bestehenden LZG-Assoziationen und neuen episodischen Assoziationen zum Erinnern von Sätzen optimal ist.

Während wir bislang nur den Hippocampus beim Abruf in einer episodischen Gedächtnisaufgabe untersucht hatten (Kuchinke et al., 2013), ist mittlerweile eine erste Pilotstudie entstanden, welche das Zusammenspiel des AROMs – als ein Modell des kortikalen LZG – mit den episodischen Prozessen in den hippokampalen Sub-Regionen beim Enkodieren adressiert (Klein, 2018). In dieser Master-Arbeit wurde ein Teil der in Roelke, Franke et al. (2018) präsentierten Prime-Zielwort-Paare im fMRT-Scanner als Paarassoziationen gelernt. Das heißt, auch in dieser Arbeit wurden die experimentellen Faktoren direkter und gemeinsamer Assoziationen experimentell manipuliert (vgl. Kapitel. 4.4). In der Verhaltensvorstudie zeigten sich Haupteffekte direkter und gemeinsamer Assoziationen mit *Cued-recall*-Raten um die 30 % für Wortpaare ohne direkte und gemeinsame Assoziationen

[23] Geschäftszeichen der Deutschen Forschungsgemeinschaft HO5139/2-1 & 2-2.

und etwa 75 % für Wörter mit direkten und gemeinsamen Assoziationen (Abb. 17, Klein, 2018, p. 40). Die Idee des fMRT-Experimentes war, dass die über das AROM definierten (kortikalen) LZG-Informationen in das hippokampale System einfließen (Rolls, 2007). Hier werden zunächst neue, rein episodische Wissenskombinationen im DG entdeckt (vgl. Kumaran & McClelland, 2012; Rolls, 2007). Dies konnte Herr Klein (2018) bestätigen: Es fand sich während dem Lernen ein Haupteffekt der AS in einer Region (nucleus caudatus; vgl. Abb. 19, Klein, 2018, p. 45), die er dem DG zuordnete (p. 50): Wortpaare ohne direkte Assoziationen lösten höhere Aktivierungen in dieser Region aus. Über den perforanten Pfad umgeht die in den Hippocampus einfließende Information zum Teil auch den DG (Abb. 2, Yassa & Stark, 2011) und fließt direkt in das sehr stark rekurrent verschaltete Subfeld 3 im cornu ammonis (CA3). Herr Klein fand einen Interaktionseffekt zwischen direkten und gemeinsamen Assoziationen in einer Region (gyrus parahippocampalis; Abb. 20, Klein, 2018, p. 46), die sich wahrscheinlich der CA3-Region zuordnen ließ (p. 50). Diese Interaktion konnte durch einen Effekt gemeinsamer Assoziationen innerhalb der Wörter ohne direkte Assoziation erklärt werden (p. 46). Ähnlich wie in rekurrent verschalteten neuronalen Netzwerkmodellen (z. B. Elman, 1990) wird in der ebenso stark rekurrent verschalteten CA3-Region wahrscheinlich ein Prozess angestoßen, der es erlaubt, über viele gemeinsame Assoziierte zu generalisieren (vgl. Franke et al., 2017).

Wie ließe sich das Gesetz des Kontrastes symbolisch simulieren?

Wenn man sich fragt, welche Prozesse in dem symbolisch repräsentierenden AROM noch nicht abgebildet werden, und einen genauen Blick in Jeffrey Elmans (1990) Artikel „*Finding structure in time*" wirft, dann ergibt sich als mögliche Antwort, dass SRN-Modelle eine sogenannte *Exclusive-Or*-(XOR-)Unterscheidung abbilden können (Elman, 1990, p. 185f). Die XOR-Funktion reduziert einen 2-bit Vektor auf einen 1-bit Vektor: Wenn die beiden Bits des ersten Vektors unterschiedliche Werte haben, also 01 oder 10 sind, dann ist der Output-Vektor 1. Er ist 0, wenn beide Einträge denselben Wert aufweisen (11 oder 00). Wie ließe sich damit das Assoziationsgesetz des Kontrastes abbilden? Wenn wir zwei Wörter auf deren Assoziation zu einem potentiell gemeinsamen Assoziierten untersuchen, dann würde eine XOR-Funktion das Detektieren von distinkten, semantischen Eigenschaften erlauben (vgl. Gamallo, 2018). Die XOR-Funktion wäre 1, wenn ein Wort mit dieser Eigenschaft assoziiert ist, das andere aber

nicht. Eine aktuelle Arbeitshypothese lautet, dass diese Funktion es erlaubt, die hippokampale Funktion der Mustertrennung im DG abzubilden (z. B. Kumaran & McClelland, 2012). Da so eine möglichst trennscharfe Repräsentation von zu erinnernden Assoziationen erzeugt wird, ließe sich dies als eine operationale Definition des Assoziationsgesetzes des Kontrastes betrachten (vgl. Kapitel 6.2, Hofmann & Kuchinke, 2015).

Während die gemeinsamen Assoziierten dazu genutzt werden können, Synonyme zu entdecken (Rapp, 2002), würde die automatisierte Suche nach Antonymen möglicherweise von dieser operationalen Definition des Gesetzes des Kontrastes profitieren. Denn auf der einen Seite haben Antonyme viele gemeinsame semantische Eigenschaften: So haben zum Beispiel die Wörter „Heirat" und „Scheidung" gemeinsame Assoziierte, wie *Ehe* und *Frau*. Dies spricht auf der einen Seite dafür, dass das Gesetz der Ähnlichkeit durch die Anzahl gemeinsamer Assoziierter teilweise abbildbar wäre. Auf der anderen Seite existieren Wörter, die nur mit einem der beiden Wörter gemeinsam auftreten. Zum Beispiel hat das Wort „Heirat" semantische Eigenschaften, wie *Paar* und *Geburt*, die nicht mit „Scheidung" assoziiert sind. „Scheidung" hingegen ist mit *Sorgerecht* und *Trennung* assoziiert, nicht jedoch mit „Heirat". Solche distinkten semantischen Eigenschaften ließen sich durch eine Funktion im Hippocampus detektieren, die die trennscharfen Repräsentationen dann in das LZG überträgt, um dieses Wort von auf anderen „semantischen Dimensionen" ähnlichen Wörtern unterscheiden zu können.

6.3. Intelligente Lernprozesse in Algorithmen

In der Einleitung in Kapitel 1 haben wir die Herausforderung aufgeworfen, präquantitative Theoriebildung für die allgemeine Psychologie zu überwinden. Wir haben uns gefragt, wie weit uns verbale Theorien zum Beispiel für diagnostisch relevante Fragestellungen bringen können. Hier möchte ich eine mögliche, algorithmisch informierte Antwort auf die alte Frage der Psychologie aufgreifen: Was ist Intelligenz? Wie Amelang und Bartussek (1997) zusammenfassen, „*haben verbale Definitionen keinen substantiellen Beitrag zum Verständnis und der Erforschung des Konstruktes Intelligenz leisten können*" (p. 189). Damit bietet die operationale Definition eine für die psychologische Forschung derzeit besser quantifizierbare Antwort auf die Frage, was Intelligenz ist: Intelligenz ist das, was die Tests messen (vgl. Amelang & Bartussek, 1997; Boring, 1923).

In Kapitel 4.2 verwenden wir ein RNN, um Varianz an einer Lückentextergänzungsaufgabe aufzuklären (Hofmann et al., 2017). Eine andere Aufgabe, auf die man RNNs bereits trainieren konnte, ist das Beantworten von Analogie-Fragen (Mikolov, Yih, & Zweig, 2013): Was der König für die Königin ist, ist der Onkel für die ...? Die richtige Antwort wäre „Tante" (Jurgens, Turney, Mohammad, & Holyoak, 2012). Ein dritter Typus Aufgabe, den solche Modelle lösen konnten, sind Synonym-Aufgaben (Landauer & Dumais, 1997; Rapp, 2002; Zhila, Yih, Meek, Zweig, & Mikolov, 2013). Da diese drei Aufgaben in ähnlicher Form bereits in den Subtests Satzergänzung, Analogien und Gemeinsamkeiten des Intelligenz-Struktur-Tests 2000R enthalten waren (IST 2000R, Liepmann, Beauducel, Brocke, & Amthauer, 2007; vgl. auch Kersting, Althoff, & Jäger, 2008), sollten solche AI-Modelle wahrscheinlich ein gewisses Maß an „Intelligenz" aufweisen.

Wenn RNNs Intelligenzaufgaben ähnlich wie Menschen lösen können, wird es möglich, die kognitiven Prozesse beim Wissenserwerb besser zu verstehen und damit hinreichende Bedingungen zu spezifizieren, die zu intelligentem Verhalten führen (vgl. Simon & Newell, 1971, p. 146). Wir können aus verschiedenen Algorithmen verschiedene Definitionen des Prozesses ableiten, der zu mehr oder weniger intelligent verwendbaren Wissensstrukturen führt. Eine verbaltheoretische teilweise Umschreibung dessen was „Intelligenz" gemäß RNNs wäre, könnte vielleicht so lauten: Dieser Algorithmus, der zu intelligentem Verhalten führt, verfügt über die Fähigkeit, aus der Erfahrung mit der Umwelt „abstrakte" Wissenseinheiten zu vernetzen, um eine Vielzahl unterschiedlicher Beobachtungen durch weniger *hidden units*

abzubilden. Als verbaltheoretische Zusammenfassung klingt diese Definition freilich sehr vage. Trainieren wir jedoch das RNN an einem Korpus, können wir diese intelligente Struktur berechnen und damit schätzen welches „intelligente" Verhalten aus der konkreteren formalen Definition folgt.

Eine Möglichkeit, das relative Ausmaß von Intelligenz zu definieren, ist die Übereinstimmung mit einer Normstichprobe. Wenn wir die Versuchspersonen, welche die empirischen LTEW aus Kapitel 4.2 generiert haben, als Normstichprobe begreifen, dann gibt die Varianzaufklärung der Sprachmodelle an, wie gut das Modell in der Lage ist, das intelligente Verhalten von Menschen in Lückentextergänzungsaufgaben zu imitieren.

Interessanterweise klärt das *N-gram*-Modell in allen Fällen deskriptiv mehr Varianz an den empirischen LTEW auf die von uns verwendeten SRNs (Tabelle 10.2, Hofmann et al., 2017; Mikolov, 2012). Was wäre „*Intelligenz, wie sie unsere Tests messen*" (Boring, 1923) gemäß der Definition des *N-gram*-Modells? Für ein *N-gram*-Modell werden einfache Häufigkeiten ausgezählt, dass ein Wort N auftritt, gegeben die vorangegangenen Wörter. Hieraus werden bedingte Wahrscheinlichkeiten für alle Wörter errechnet. Wir könnten damit die These stützen, dass einfache bedingte Wahrscheinlichkeiten, das heißt positionsspezifische, frequenzabhängige Kontiguität, das menschliche Verhalten besser abbildet als die oben genannte, alternative Definition von „Intelligenz" gemäß unseren SRNs.

Vergleichen wir die beiden Definitionen von RNNs und dem *N-gram*-Modell mit wegweisenden Definitionen aus der Geschichte der Intelligenzforschung (Catell, 1943, p. 178[24]): „*Fluide Fähigkeit hat den Charakter einer rein allgemeinen Fähigkeit, Relationen zwischen fundamentalen Eigenschaften wahrzunehmen und zu diskriminieren*"; „*Kristalline Fähigkeit besteht aus Unterscheidungsgewohnheiten, die sich in einem spezifischen Feld seit langem ausgebildet haben*". Da sich bedingte Wahrscheinlichkeiten als operationale Definition von Unterscheidungsgewohnheiten bezüglich spezifischer Wörter verstehen lassen, könnte man das *N-gram*-Modell näher an den Begriff der kristallinen Intelligenz rücken. Im Gegensatz dazu kann man zeigen, dass sogar noch einfachere neuronale Netzwerkmodelle als das SRN generalisieren können – eine Fähigkeit, die man der fluiden Intelligenz zuschreibt (z. B. Klauer,

[24] „Fluid ability has the character of a purely general ability to discriminate and perceive relations between any fundaments". „Crystallized ability consists of discriminatory habits long established in a particular field".

Willmes, & Phye, 2002). Generiert man eine neue Input-Einheit, wie zum Beispiel den Namen eines neuen, unbekannten Vogels, und trainiert es auf einige wenige bekannte Kontexte, wie zum Beispiel „kann *fliegen*" und „hat *Flügel*", dann lernt das Netzwerk sehr schnell, diesem neuen Begriff ein ähnliches *Hidden-unit*-Aktivierungsmuster aus „*fundamentalen Eigenschaften*" zuzuweisen (Catell, 1943, p. 178); dies erlaubt die Vorhersage neuer konkreter Eigenschaften, wie zum Beispiel „kann *wachsen*", die das Netzwerk so nicht trainiert hat (Rumelhart & Todd, 1993, zitiert aus McClelland & Rogers, 2003, p. 314). Weil neuronale Netzwerkmodelle, wie SRNs, generalisieren können, sind sie wahrscheinlich auch in der Lage fluide Aspekte von Intelligenzleistungen abbilden.

Folglich lässt sich festhalten, dass diese Algorithmen unsere Intelligenzaufgabe mit zwei verschiedenen Strategien lösen können: Entweder der Algorithmus hat ein sehr genaues Gedächtnis und nutzt die bedingte Wahrscheinlichkeit, dass ein Wort N auftritt, gegeben eine vorangegangene Zeichenfolge aus N-1 Wörtern (*N-gram*-Modell), oder er kann „latente Dimensionen" aus der Lerngeschichte erschließen, so dass das Modell darauf trainiert werden kann, das jeweils nächste Wort vorherzusagen (SRN).

Dieser Unterschied – das stärkere Abbilden fluider Intelligenzleistungen in SRNs und die eher kristalline Intelligenz in *N-gram*-Modellen – lässt sich an der Nutzung komputationaler Ressourcen der beiden Algorithmen aufzeigen: Während ein *N-gram*-Modell in der Regel einen höheren Festplattenspeicher benötigt, verbraucht ein modernes SRN in der Regel viel mehr Rechenzeit, um zu einer sparsam mit dem Speicherplatz umgehenden, „abstrakten" Wissensstruktur zu gelangen. Das *N-gram*-Modell löst die Intelligenzaufgabe primär mit dem „Gedächtnis", während ein SRN den Speicherplatz durch rechenintensivere Inferenzen und Abstraktionsleistungen, wie Generalisierung, verkleinern kann (McClelland & Rogers, 2003). Dass auch Informationen über die relative „Abstraktheit" der Wörter in solchen Algorithmen enthalten sind, lässt sich auch daran zeigen, dass sie *concreteness ratings* vorhersagen können (Westbury et al., 2013; Hollis et al., 2017; vgl. Kapitel 4.1), da „Abstraktheit" und „Konkretheit" häufig als Gegenpole einer semantischen Dimension betrachtet werden (z. B. Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011).

Die Vorhersage der SFD ist zwar sicherlich kein allgemein anerkanntes Kriterium für „Intelligenz". Dennoch möchte ich für einen Moment mit der Hypothese arbeiten, dass die Vorhersage dieses Blickbewegungsmaßes eine andere Form intelligenten Verhaltens abbildet,

beispielsweise die Simulation eines anderen kognitiven Systems im Sinne der *theory of mind* – diese Simulation erlaubt die Vorhersage des Verhaltens einer anderen Person (z. B. Gallese & Goldman, 1998; Wimmer & Perner, 1983). Beim SRN sehen wir auf Basis eines kleineren, aber für die gesprochene Sprache repräsentativeren Trainings-Korpus deskriptiv etwas bessere SFD-Vorhersagen als beim *N-gram*-Modell (Tab. 10.4, Hofmann et al., 2017; vgl. Brysbaert et al., 2011), obgleich dieses deutlich kleinere *Subtitles*-Korpus naturgemäß mehr Rauschen enthalten sollte als die größeren Korpora. Wenn wir im *N-gram-Modell* den kristallinen Gedächtnisaspekt guter Intelligenz-Leistungen und im SRN die Fähigkeit zu abstrahieren abgebildet sehen, könnten wir hieraus die Schlussfolgerung ziehen, dass die Fähigkeit zu abstrahieren auch bei einer kleineren, aber vermeintlich repräsentativeren Erfahrungsgrundlage das menschliche Blickbewegungsverhalten gut abbildet (vgl. Brysbaert et al., 2011). Wird auf diese Weise gelernt, wäre eine geringere Menge unterschiedlicher Erfahrungen ausreichend. Bei der gedächtnisintensiveren Strategie des *N-gram*-Modells ist die größere Erfahrungsgrundlage des *News*-Korpus für die Vorhersage von LTEW hingegen häufig hilfreich (Tabelle 10.2, Hofmann et al., 2017).

Zwar ist es durch die Konkretisierung in Algorithmen möglich, den Prozess der Intelligenzbildung besser zu verstehen als dies auf Grund einer rein verbalen Beschreibung alleine möglich wäre. Aber das bedeutet nicht, dass das Verständnis für diese Algorithmen einfacher wäre als das Verständnis durch die klassischen, von der Intuition der Konstrukteure und der Normstichprobe geprägten Intelligenztests. Gerade im Vergleich zu dem relativ einfachen SRN von Jeffrey Elman (1990) offenbaren moderne Weiterentwicklungen, wie *word2vec* (Mikolov, Chen et al., 2013), immer mehr komputationale Details, deren spezifische Vorteile für die Vorhersage von Sprachdaten auch von den Experten erst nach und nach verstanden werden müssen (Levy et al., 2015). Laut Feynman ist es jedoch „*nicht wichtig wo die Schätzung herkommt. Es ist nur wichtig, dass sie mit dem Experiment übereinstimmen sollte*" (Seabala, 2018, 6:57[25]).

In der Zukunft gilt es festzustellen, welchen partiellen Intelligenzquotienten man bei solchen Algorithmen in tatsächlichen Intelligenz-Subtests messen kann (z. B. Liepmann et al., 2007). Darüber hinaus wäre zu fragen, wie die Praxis der Intelligenz-Testung von einem algorithmisch

[25] „[It's not] important where the guess comes from, it's only important that it should agree with [the] experiment"

konkretisierten theoretischen Verständnis profitieren könnte. Eine immer wieder aufkommende Kritik an Intelligenztests ist, dass diese eine spezifische Erfahrungsgrundlage voraussetzen. Deshalb sieht bereits Catell (1943, p. 157[26]) die Herausforderung darin, diese Tests „*von Annahmen gleichförmigen Wissens (...) über alle Erwachsenen zu befreien*". Dass Intelligenztests auf Basis einer Erfahrungsgrundlage generiert wurden, die für einen mittleren bis hohen sozioökonomischen Status typisch ist, könnte einer der Gründe dafür sein, warum Kinder mit einem niedrigen sozioökonomischen Status in Intelligenztests typischerweise schlechter abschneiden (z. B. Stumm & Plomin, 2015). Wir können ein *N-gram-Modell* oder ein RNN an unterschiedlichen Korpora trainieren. Diese beiden Modelle werden unterschiedliche bedingte Wahrscheinlichkeiten lernen oder die *hidden units* unterschiedlich mit Wörtern verknüpfen. Hat ein Kind ein gutes Gedächtnis, würden wir bessere Vorhersageleistungen von einem *N-gram-Modell* erwarten. Kann es eher abstrahieren, sollte ein RNN mehr Varianz an seinen Leistungen aufklären. Bei Kindern mit einer ausgeprägten Fähigkeit zu abstrahieren könnte es für den Wissenserwerb nützlicher sein, weniger Textmaterial zu lernen. Dieses Material sollte jedoch von Anfang an alle notwendigen Informationen enthalten, aus welchen sie alle „latenten semantischen Dimensionen" über die *hidden units* lernen können (Mikolov, 2012; vgl. Tabelle 10.4, Hofmann et al., 2017).

Wenn man ein RNN als allgemeine Fähigkeit begreift, Relationen zwischen fundamentalen Prinzipien zu erkennen, die es erlauben, wahrscheinlich auftretende, zukünftige Ereignisse vorwegzunehmen und zwischen vielen Objekten mit wenigen Prinzipien zu unterscheiden (vgl. Catell, 1943), dann entspricht dies sicherlich einer algorithmischen Definition relativ fluider Intelligenz-Fähigkeiten. Diese allgemeine Fähigkeit könnten wir abbilden, indem wir die Sprachmodelle an einem Korpus trainieren, das die typische Erfahrungsgrundlage von Individuen mit einem niedrigen sozioökonomischen Status darstellt. Auf Basis eines besser passenden Korpus, das für das spezifische Wissen dieser Personen repräsentativ ist, ließe sich wahrscheinlich ein wesentlich fairerer Intelligenztest errechnen. Auf Grund der unterschiedlichen Erfahrung werden diese Individuen sicher andere Prinzipien abstrahieren. Eine andere Personengruppe, für welche die relativ geringen Leistungen in Intelligenztests nur bedingt die Schlussfolgerung auf eine geringe Intelligenz erlauben, sind ADHD-Kinder. Diese Kinder zeigen insbesondere dann gute Leistungen, wenn sie hochgradig motiviert sind (z. B.

[26] „The greater need for freeing adult tests from assumptions of uniform knowledge, education, and skills".

McInerney & Kerns, 2003). Wenn wir wissen, wofür sich ein solches Kind interessiert, wäre es möglich, spezifische Intelligenztests für solche Kinder zu entwickeln. Eine aus der selbstgewählten Erfahrungsgrundlage dieser Kinder errechnete Lückentextergänzungsaufgabe würde die Wahrscheinlichkeit erhöhen, dass sie diese Aufgabe aufmerksam bearbeiten, und damit eher eine Maximalleistung auf Basis der eigenen Interessen anzeigen. Wenn es Wörter einsetzt, die überwiegend auf Basis eines RNNs vorhersagbar sind, würde es darüber hinaus von einer abstraktionsbasierten Lernstrategie besser profitieren. Ein möglicher therapeutischer Ansatz für ADHD-Kinder wäre, deren Interessensspektrum zu erweitern, indem man diesen Texte präsentiert, die sich – ausgehend vom Interessensbereich des Kindes – immer mehr den zu vermittelnden Inhalten systematisch annähern. Über semantisch ähnliche Zwischenschritte zu den Ziel-Inhalten ließe sich in ADHD-Kindern ein breiter gefächertes Interesse erzeugen, welches sie in die Lage versetzt, immer mehr Inhalte mit Aufmerksamkeit zu versehen. Ich habe bereits eine Abschlussarbeit betreut, in der die Versuchspersonen zwei Monate auf einem zur Verfügung gestellten Tablet einem möglichst natürlichen Leseverhalten nachgehen sollten (Pasche, 2018). Indem man aus den so erhobenen Inhalten ein individuelles Korpus erzeugt, sollte es möglich werden, Sprachmodelle auf Basis der individuellen Erfahrungsgrundlage eines Menschen zu trainieren, um damit individuelle semantische Strukturen abzubilden und diese perspektivisch für Diagnostik und Therapie einzusetzen.

7. Schlussfolgerungen

Die Geschichte der Forschung zu den Assoziationsgesetzen der Kontiguität, Frequenz, Ähnlichkeit und des Kontrastes lässt sich in zwei Phasen einteilen: Bis zur kognitiven Wende in den 50er Jahren des 20. Jahrhunderts erklärten Psychologen, wie Ebbinghaus, Jung, Skinner und Hebb, ihre Daten rein verbaltheoretisch. Ab der kognitiven Wende machten es Ansätze, wie die Informationstheorie, verteilt lernende neuronale Netzwerkmodelle, vollständig symbolisch repräsentierende Modelle der visuellen Worterkennung und die Analyse großer Textkorpora, möglich, so komplexe Prozesse, wie die Entstehung und Verwendung semantischer LZG-Strukturen, in Algorithmen quantitativ zu beschreiben (Kapitel 1).

In Kapitel 2 präsentieren wir das IAM und seine Erweiterung um eine Ebene der Semantik im AROM und stellen die Frage, wie solche Modelle mit neurokognitiven Daten evaluiert und weiterentwickelt werden können (Hofmann & Jacobs, 2014; vgl. Kapitel 5.1; Jacobs & Hofmann, 2013). Auf Grund der Diskussion von Verhaltens-, EEG- und fMRT-Effekten der Länge, Sequenz, Frequenz, Wiederholung, orthographischen Nachbarschaft und „Semantik" (vgl. auch Tabelle 1 in Radach & Hofmann, 2016; Kapitel 5.2) schlagen wir eine Reihe von *model-to-data connections* vor, welche die Frage beantworten sollen, wie IAMs mit neurokognitiven Daten geprüft werden können. Visuelle Informationen erreichen *bottom-up* getrieben den OC etwa ab 60 ms. Sie lassen sich durch die Ebene visueller Eigenschaften abbilden. Frühestens ab 100 ms trifft die Information auf den FFG als das visuelle Wortformareal (Cohen et al., 2001; Sereno et al., 1998), und der durch die Wortfrequenz vermittelte lexikalische Zugriff auf ein orthographisches Lexikon findet statt (z. B. Fiebach, Friederici, & Cramon, 1999). Wenn man einen simulierten AROM-Modellzyklus mit 50 ms im realen Experiment gleichsetzt (Kapitel 6.1), erreichen *Bottom-up*-Informationen die Ebene der Semantik ab 200 ms (Kutas & Federmeier, 2011). Da sie dort erst noch mit den Assoziationen wahrscheinlich vorher präsentierter Wörter interagieren müssen, erklärt sich, warum N400-Effekte der semantischen Integration im Temporallappen und semantischen Kompetition im IFG häufig in späteren Zeitfenstern auftreten (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997). Weil sich Negativierungen auf dieser Komponente durch lexikalische Kompetition erklären lassen, können wir die N400 als einen Spezialfall einer N2 begreifen, die wahrscheinlich aus Aktivierungen im ACC resultiert (Yeung et al., 2004).

Mit einem IAM ließen sich jedoch nur die N2-/N400-Effekte weitgehend bedeutungsfreier Stimuli, wie zum Beispiel Nichtwörter, simulieren (Braun et al., 2006; Hofmann, Tamm et al.,

2008). Ebenso fanden Binder und Kollegen (2003) bei Wörtern mit einer geringen Anzahl orthographischer Nachbarn kompensatorische semantische Aktivierungen im Temporallappen. Solche EKP- und fMRT-Ergebnisse legten nahe, dass IAMs um eine vollständig implementierte semantische Ebene erweitert werden sollten. Klassische operationale Definitionen von „semantischen" Variablen für Worterkennungs- und Satzverarbeitungsaufgaben basierten auf der freien Assoziationsaufgabe und der LTEW (Kutas & Hillyard, 1984; Lucas, 2000). Diese Definitionen haben den Nachteil, dass vom menschlichen Verhalten zirkulär auf anderes menschliches Verhalten geschlossen wird. Darüber hinaus eignen sich solche Definitionen nur dafür, die stärksten Assoziationen zwischen Wörtern zu definieren (McKoon & Ratcliff, 1992). Ebenso lässt sich eine allgemeingültige operationale Definition von „Semantik" auf Grund der gigantischen Menge an durchzuführenden Vorexperimenten nur schwer realisieren (vgl. Hofmann et al., 2018, p. 2288). Eine einfache Lösung für diese Probleme bietet das kookkurrente Auftreten von Wörtern in den Sätzen von großen Korpora, welches die Grundlage für die AS im AROM darstellt (Dunning, 1993; Hofmann et al., 2011). Definiert man damit die assoziativen Relationen der Wörter einer episodischen Gedächtnisaufgabe, lässt sich die assoziative Aktivierungsausbreitung über ko-aktivierte andere Stimuli dynamisch simulieren und zur Vorhersage von Erinnerungsleistungen bei nicht-studierten und studierten Wörtern einsetzen (Hofmann et al., 2011; Stuellein et al., 2016). So kann man Fehlerinnerungen bei nicht-studierten Wörtern und die Verbesserung von Erinnerungsleistungen für studierte Wörter auf spezifische Assoziationen innerhalb des Reizmaterials zurückführen (Kimball et al., 2007; Roediger & McDermott, 1995). Ebenso führt die Annahme höherer Ruheaktivierungen bei studierten Wörtern ohne weitere Vorannahmen zu einer höheren Gedächtnissignalvarianz als bei nicht-studierten Wörtern (Glanzer et al., 1999). Der Erklärungswert des AROMs lässt sich auch mit einer Re-Analyse von fMRT-Daten steigern (Forgács et al., 2012): Je stärker die AS zwischen den Nomina eines Kompositums, desto geringer ist die semantische Kompetition und damit auch die Aktivierung des IFG.

Das Experiment in Kapitel 3 basiert auf einer klassische Definition „semantischer" Zusammenhänge, der LTEW: Es werden OC- und OFC-Aktivierungen mittels fNIRS untersucht (Hofmann et al., 2014). Wenn die Zielwörter im Satz hochgradig vorhersehbar sind, dann lösen sie geringere hämodynamische Antworten im OC aus als unvorhersehbare Wörter. Dies lässt sich mit voraktivierten visuellen Eigenschaften bei vorhersehbaren Wörtern erklären (vgl. Kapitel 5.2; Abb. 2, Radach & Hofmann, 2016). Darüber hinaus zeigen wir, dass bei seltenen Wörtern der OFC bei gleichzeitig geringer Wortvorhersagbarkeit aktiviert wird. Doch

auch die hochfrequenten Wörter scheinen einen Beitrag zu diesem Interaktionseffekt zu leisten. Bei diesen zeigt sich eine nicht-signifikante Tendenz in Richtung des umgekehrten Effektes der Wortvorhersehbarkeit: Es findet sich eine höhere OFC-Aktivierung für hochgradig vorhersehbare Wörter. Somit lassen sich die widersprüchlichen Vorhersagen bezüglich der Rolle des OFC im Umgang mit Erwartungs-Bestätigung (Bar et al., 2006) oder Erwartungs-Verletzung (Nobre et al., 1999) auflösen: Bei niederfrequenten Wörtern reicht die *Bottom-up*-Information innerhalb von 250 ms nicht aus, um den Stimulus eindeutig zu identifizieren. Deshalb reagiert der OFC auf gleichzeitig unvorhersehbare Ereignisse. Wenn jedoch eine hohe Menge an *Bottom-up*-Informationen bei hochfrequenten Wörtern verfügbar ist, dann findet sich eine Tendenz, auf vorhersagbare Wörter stärker zu reagieren. Da dieses Bestätigungssignal der Vorhersage bei hinreichend vielen *Bottom-up*-Informationen nur dann zu erwarten war, wenn die Versuchspersonen mit einer *top-down* getriebenen Strategie lesen (Bar et al., 2006; Dambacher et al., 2012), präsentieren wir die Sätze, welche die Zielwörter enthalten, mit einer lesenahen Präsentationsrate von 250 ms – mit 30 ms weißem Bildschirm zwischen den Wörtern, um die Verringerung visueller Informationsaufnahme während einer Sakkade anzunähern. Durch die relativ hohe *Sampling*-Frequenz lassen sich die hämodynamischen Antworten auf ultraschnelle Ereignisfolgen, wie Wörter bei leseähnlichem Verarbeitungstempo, voneinander differenzieren. So wird es möglich, die empirischen hämodynamischen Antworten der Zielwörter mit einer Dekonvolutionsanalyse von den umliegenden Wörtern im Satz getrennt zu betrachten (Abb. 6 & 7, Hofmann et al., 2014). Diese Arbeit lässt sich als Wegbereiter für eine Studie begreifen, in der wir die hämodynamischen Antworten beim natürlichen Lesen mit gleichzeitiger Blickbewegungsmessung untersucht haben (Roelke, Hofmann et al., 2018): Hier haben wir gezeigt, dass sich die OFC-Interaktion zwischen Wortfrequenz und Wortvorhersehbarkeit beim natürlichen Lesen in den OC verlagern kann: Für das Wort, das sich rechts vom aktuell fixierten Wort befand, zeigen sich retinotope Effekte im linken OC. Damit wurde die Methode der frNIRS in die Leseforschung eingeführt. Die fehlenden OFC-Effekte in dieser frNIRS-Studie zeigten, dass sich diese in Hofmann et al. (2014) durch die Menge verfügbarer *Bottom-up*- und *Top-down*-Informationen in einer vorgegebenen Zeit erklären lassen (z. B. Dambacher et al., 2012).

Doch wenn die abhängige Variable aus einem Vorexperiment, wie die LTEW, zur unabhängigen Variable in einem anderen Experiment gemacht wird, ist letztere per se nicht unabhängig vom menschlichen Verhalten (z. B. Kapitel 2, Hofmann & Jacobs, 2014). Genau genommen können wir einen solchen klassischen Ansatz als multivariates Vorgehen begreifen, das heißt, es werden zwei von der menschlichen Performanz abhängige Variablen miteinander

verglichen, die aus unterschiedlichen Personenstichproben stammen (Hofmann et al., 2018). Diese Zirkularität gilt es zu vermeiden, um damit neue Erklärungsebenen für die Psychologie zu erschließen (Hempel & Oppenheim, 1948) und das „semantische" System des Menschen aus seiner Erfahrung abzuschätzen (Hofmann et al., 2018; Hofmann & Jacobs, 2014; Westbury, 2016).

In Kapitel 4 werden unterschiedliche algorithmische Ansätze präsentiert, die demonstrieren, wie diese Zirkularität umgangen werden kann. Im Allgemeinen werden diese Modelle an einem Korpus trainiert, das die Erfahrungsgrundlage eines Menschen annähernd abbilden soll. Zwar wird diese in Kapitel 4 noch über Standard-Korpora, zum Beispiel Wikipedia, abgebildet, aber es wurde bereits eine Pilotstudie durchgeführt, in der das möglichst natürliche Leseverhalten zweier Versuchspersonen über zwei Monate dazu verwendet wurde, individuelle Korpora zu erzeugen (Pasche, 2018; vgl. Kapitel 6.3). Solche individuellen Korpora sollten auch den großen Internet-Firmen zur Verfügung stehen. Mit diesen individuellen Korpora ließe sich deren Repräsentativität, zum Beispiel für die Interessen der Versuchspersonen, abschätzen, wenn wir die Ergebnisse der resultierenden Sprachmodelle zum Beispiel mit den Daten der generellen Interessensskala vergleichen (Brickenkamp, 1990). So können wir die Möglichkeiten und Grenzen von *big data* aufzeigen (vgl. Pasche, 2018). Dies ist eine Perspektive, die einen empirischen Beitrag zur Diskussion über gesellschaftliche Fragen, wie zum Beispiel Vorratsdatenspeicherung, bieten soll. Nachdem die Simulation der Gedächtniskonsolidierung auf Basis solcher Korpora abgeschlossen ist, sollen solche Sprachmodelle Vorhersagen über das menschliche Verhalten ermöglichen.

Wir nutzen in Kapitel 4.1 verschiedene computerlinguistische Ansätze, um damit das Verhalten in Experimenten vorherzusagen, wie sie typischerweise als Vorstudien für Priming-Experimente verwendet werden (z. B. Dimigen, Kliegl, & Sommer, 2012). Durch algorithmische Modelle sollte Varianz aus je 300 Assoziations-*ratings* in insgesamt drei Studien aufgeklärt werden (Hofmann et al., 2018). Die meisten algorithmischen Modelle des semantischen Gedächtnisses in der Psychologie beschreiben die Bedeutung eines Wortes verteilt auf subsymbolische Einheiten (z. B. Bhatia, 2017; Elman, 1990; Griffiths et al., 2007; Landauer & Dumais, 1997; Mandera et al., 2017; Mikolov, 2012). Das *Skip-gram*-Modell, als ein solcher subsymbolischer Ansatz (Mikolov, Chen et al., 2013), erklärt reproduzierbar etwa 50 % der *Item-level*-Varianz der Assoziations-*ratings*. In unseren Analysen mit *linear mixed effects models* konnte aber auch unsere symbolisch repräsentierende Prädiktorvariable, die AS,

einen signifikanten Varianzanteil erklären. Als Einzelprädiktor klärt die AS jeweils etwa 40 % der *Item-level*-Varianz an Assoziations-*ratings* in den drei Studien auf.

Da die AS auf einem Signifikanztest basiert (Dunning, 1993), der das überzufällig häufig gemeinsame Auftreten von Wörtern feststellt, verbindet sie die Assoziationsgesetze der Frequenz und Kontiguität. Lassen sich diese beiden Gesetze auch getrennt betrachten? Für ein reines Gesetz der Kontiguität, das von der gemeinsamen und einzelnen Frequenz der Wörter unabhängig wäre, könnten wir annehmen, dass bereits beim einmaligen gemeinsamen Auftreten zweier Reize eine Assoziation entsteht. *One-trial learning* lässt sich innerhalb der CLS-Theorie nur dann mit einem rein kortikalen LZG-Speicher erklären (Marr, 1970, 1971; McClelland et al., 1995), wenn diese Assoziation sehr Schema-konsistent ist, das heißt, wenn bereits sehr viele sehr ähnliche assoziative Strukturen gespeichert sind (McClelland, 2013). Alle komplett neuen, Schema-inkonsistenten Assoziationen würden innerhalb der CLS-Theorie auf Basis eines Lerndurchgangs nur dann gelernt, wenn ein neues *conjunction unit* entsteht – ein Prozess, der im DG des Hippocampus zu verorten wäre (Kumaran & McClelland, 2012; vgl. Kapitel 6.2). Durch *generalized replay* wird diese Information später wieder aufgerufen, und so trainiert der DG das kortikale LZG. Verfügt ein Patient aber über keinen Hippocampus, dann kann er nur kortikal lernen. Deshalb würde er eine wesentlich höhere Anzahl an Lerndurchgängen benötigen (z. B. Squire & Wixted, 2011, p. 266), und das Assoziationsgesetz der Frequenz nähme einen wesentlich höheren Stellenwert ein. Das AROM wurde als ein Modell dieses reinen LZG-Speichers konzipiert (z. B. Hofmann & Jacobs, 2014). Deshalb macht es hier Sinn, diese beiden Assoziationsgesetze gemeinsam in der *log likelihood* abzubilden (Dunning, 1993; Hofmann et al., 2011): Wenn zwei Wörter signifikant häufiger gemeinsam in Sätzen auftreten als durch deren Einzelwortfrequenz zu erwarten wäre, definieren wir sie als „assoziiert" (vgl. auch Dunning, 1993; Evert, 2005). Diese Information, ob assoziiert oder nicht, können wir als dummy-kodierte Variable betrachten (0/1). So können wir zum Beispiel die Anzahl an Assoziationen eines Stimulus zu den anderen Stimuli zählen und zur Vorhersage von EKP-Daten nutzen (Kapitel 4.3, Stuellein et al., 2016). Eine andere 0/1-kodierte Anwendung des Assoziationsgesetzes der Frequenz wäre das Zählen der gemeinsamen Assoziierten von Prime- und Zielwort, um damit semantische Priming-Effekte vorherzusagen (Kapitel 4.4, Roelke, Franke et al., 2018).

Über die dummy-kodierte Verwendung hinaus, bietet die AS die Möglichkeit einer Anwendung des Assoziationsgesetzes der Frequenz als kontinuierliche Prädiktorvariable: Wenn zwei Wörter signifikant häufiger gemeinsam auftreten als auf Basis der Wort-Einzelauftretenshäufigkeiten zu erwarten wäre, dann wird die relative AS durch den log10-

transformierten χ^2-Wert definiert (Hofmann et al., 2011; Hofmann & Jacobs, 2014, Kapitel 2). Diese quantitative Definition bietet über die 0/1-kodierte Assoziation hinaus einen deutlichen Varianzgewinn: Wenn wir die AS in einer Post-hoc-Analyse mit dem dummy-kodierten Prädiktor ersetzen, dann klärt dieser an den 900 Assoziations-*ratings* der drei Studien 25 % der *Item-level*-Varianz auf (Kapitel 4.1). Die auf dem χ^2-Wert basierte AS jedoch 38 % (Dunning, 1993; Hofmann et al., 2018). Die genauere Beschreibung einer Assoziation zwischen zwei Wörtern erhalten wir folglich, wenn wir eine kontinuierliche Operationalisierung des Assoziationsgesetzes der Frequenz verwenden (Hofmann et al., 2011, 2018).

Kapitel 4.2 beschäftigt sich mit der Frage, wie sich die Zirkularität vermeiden ließe, eine abhängige Variable, wie die LTEW, zur unabhängigen Variable in Satzverarbeitungsaufgaben zu machen (vgl. Kapitel 3). Ein *Topic*-Modell, das häufig in der Psychologie eingesetzt wird (z. B. Andrews, Vigliocco, & Vinson, 2009; Griffiths et al., 2007), schien dafür nur bedingt geeignet – wahrscheinlich, weil es den semantischen Konsolidierungsprozess, genauso wie die LSA, auf Basis des gemeinsamen Auftretens von Wörtern in Dokumenten abbildet (Deerwester et al., 1990; Landauer & Dumais, 1997). In Satzverarbeitungsaufgaben scheint ein solches *Long-range*- Semantik-Modell eine relativ geringe Rolle zu spielen.

Die besten Vorhersagen für LTEW lassen sich mit einem subsymbolisch repräsentierenden SRN oder einem symbolisch repräsentierenden *N-gram-Modell* erzielen. Beide bilden *Short-range*-Semantik auf Grund der vorangegangenen Wörter im Satz ab. Jedes dieser Modelle kann zusammen mit einer Baseline aus Wortfrequenz und der Position im Satz etwa 42-49 % der LTEW-Varianz aufklären. Für N400-Amplituden bietet ein *N-gram*-Modell, das auf dem sehr großen *News*-Korpus trainiert wurde, mit 16 % Varianzaufklärung deskriptiv die beste Vorhersage, während ein SRN, das auf dem deutlich kleineren *Subtitles*-Korpus trainiert wurde, mit 28 % die beste Vorhersage für SFDs macht. Eine mögliche Erklärung für die gute SRN-Modell-Performanz auf Basis dieses kleinen, aber vermeintlich repräsentativeren Korpus ließe sich durch die Fähigkeit eines SRN erklären, „latente semantische Dimensionen" zu erschließen, die nicht nur auf den gegebenen Wörtern basieren, sondern auch auf ähnlichen Wörtern (Hofmann et al., 2018, p. 2306).

Im Gegensatz zum *Topic*- und SRN-Modell, basiert das *N-gram-Modell* auf vollständig symbolischen Repräsentationen. Es operationalisiert das semantische System quasi ausschließlich auf Grund von Input-/Output-Relationen (vgl. Skinner, 1948; vs. Chomsky, 1959). Wir berechnen hier bedingte Wahrscheinlichkeiten, dass ein Zielwort, gegeben die vier vorausgehenden Wörter, auftritt. Das *N-gram-Modell* entspricht einer positionsspezifischen

Operationalisierung der Gesetze der Kontiguität und Frequenz. Es kann die Wortvorhersehbarkeit aus dem Satzkontext, wie sie auch die LTEW abbildet, zu einem guten Teil erklären. Dennoch lassen sich die Varianzaufklärungen an N400-Amplituden und SFDs durch Hinzunahme des klassischen, auf Basis menschlicher Performanz definierten Prädiktors der LTEW jeweils um 6 % und 4 % steigern. Wenn wir in Zukunft feststellen, dass es sich hierbei um reproduzierbare Varianzanteile handelt, dann würde die LTEW noch andere Prozesse in diesen Daten beschreiben, die diese algorithmischen Modelle (noch) nicht erfassen.

In Kapitel 4.3 haben wir die AS als dummy-kodierte Variable verwendet und die Anzahl der Assoziierten der Zielwörter zu den anderen Stimuli einer episodischen Gedächtnisaufgabe für die Vorhersage von P200- und N400-Amplituden benutzt (Stuellein et al., 2016). Studierte Wörter zeigen positivere P200- und N400-Amplituden als nicht-studierte Wörter. Ebenso zeigen Wörter mit vielen Assoziierten im Reizmaterial positivere Amplituden auf diesen Komponenten. Je höher die Aktivierungen auf der semantischen Ebene, desto positiver sind diese Komponenten (vgl. Hofmann et al., 2011). Wenn wir diese Befunde mit den *model-to-data connections* aus Hofmann und Jacobs (2014) vergleichen, dann lässt sich die P200 ab 150 ms mit dem Zugriff auf ein orthographisches Lexikon in Verbindung bringen. Dieser Effekt weist darauf hin, dass die vereinfachende Annahme einer *Top-down*-Verknüpfung von der semantischen zur orthographischen Ebene aufgegeben werden sollte (vgl. Abb. 1, Hofmann et al., 2011; vs. Abb. 6, Hofmann & Jacobs, 2014). Die stärkere Negativierung der N400 bei assoziativ wenig verknüpften Reizen lässt sich im Sinne einer N2 und einer größeren semantischen Kompetition erklären (siehe auch Kapitel 2 und 6.2). Neben der Replikation höherer Fehlerraten bei assoziativ verknüpften nicht-studierten und besseren Wiedererkennensraten bei assoziativ verknüpften studierten Wörtern (Hofmann et al., 2011) fanden wir hier auch jeweils langsamere und schnellere Reaktionszeiten. Diese lassen sich mit den Entscheidungsmechanismen des MROMs erklären: Starke Aktivierungen in den ersten Modellzyklen lösen einen *Fast-guess*-Mechanismus aus, der ein liberaleres Entscheidungskriterium für die Ja-Antwort setzt. Darüber hinaus führen diese frühen Aktivierungen zu einer Verlängerung der zeitlichen Deadline für die Nein-Antwort (Grainger & Jacobs, 1996). Dies zeigt, dass die im MROM für die lexikalische Entscheidungsaufgabe entworfenen Entscheidungsmechanismen ohne weiteres auch auf episodische Gedächtnisaufgaben angewendet werden können: Semantische Ko-Aktivierungen bei studierten und nicht-studierten Stimuli wirken genauso auf Reaktionszeiten wie orthographische Ko-Aktivierungen bei Wörtern und Nicht-Wörtern (Grainger & Jacobs, 1996).

Assoziative Priming-Effekte werden in klassischen Bahnungsaufgaben multivariat durch menschliche Leistungen in der freien Assoziationsaufgabe erklärt (z. B. Lucas, 2000). In Kapitel 4.4 testen wir die Vorhersage, dass die frequenzgewichtete Kontiguität der AS solche Effekte in gebahnten lexikalischen Entscheidungen erklären kann (Roelke, Franke et al., 2018). Ebenso prüfen wir, ob klassische semantische Priming-Effekte durch die Anzahl der gemeinsamen Assoziierten vorhergesagt werden können (Lucas, 2000). Die gemeinsamen Assoziierten bilden damit auch das Gesetz der Ähnlichkeit ab. Die Vorhersagen lassen sich durch die behavioralen Daten aus einem Verhaltensexperiment und einer fMRT-Studie bestätigen: Während sich assoziative Effekte sowohl bei einer SOA von 200 ms als auch bei 1000 ms antworterleichternd auswirken, bleiben die faszilitatorischen Effekte der gemeinsamen Assoziierten auf die kurze SOA beschränkt. Dies entspricht einem typischen Befundmuster in klassischen Bahnungsaufgaben (vgl. Hutchison, 2003; Lucas, 2000) und bestätigt damit die Annahme, dass assoziative und semantische Priming-Effekte jeweils mit direkten und gemeinsamen Assoziationen erklärt und simuliert werden können (Kapitel 6.1). Ebenso ermöglicht unsere operationale Definition dieser beiden Variablen durch die AS eine Antwort auf McNamaras (2005, p. 86) Herausforderung, „*auf irgendeine plausible Art und Weise zwei hochgradig assoziierte Wörter zu finden, die nicht semantisch relatiert sind*". Mit dieser vorher nur theoretisch postulierbaren experimentellen Zelle untersuchen wir assoziatives und semantisches Priming meines Wissens zum ersten Mal in einem voll faktoriellen Design (vgl. Ferrand & New, 2003). Ebenso bestätigen wir die in Kapitel 2 vorgeschlagenen *model-to-data connections* für fMRT-Daten in Roelke et al. (2016, vgl. Abb. 1B). Wörter mit vielen gemeinsamen Assoziierten lösen geringere Aktivierungen im OC, FFG, ATL und IFG aus, was eine vollständige, *Bottom-up*- und *Top-down*-Interaktivität zwischen den visuellen, orthographischen und semantischen Repräsentationsebenen nahelegt. Über viele gemeinsame Assoziierte zum Prime wird das Zielwort voraktiviert, weshalb für dessen Erkennen weniger neuronale Energie verbraucht wird. Insgesamt bieten die Ergebnisse aus den Kapiteln 3, 4.3 und 4.4 also konvergierende Evidenzen für eine vollständige *Bottom-up*- und *Top-down*-Interaktivität aller Ebenen des AROMs. Man sollte das AROM also vollständig interaktiv verdrahten, wenn hinreichend viele komputationale Ressourcen zur Verfügung stehen, um ein deutlich komplexeres, revidiertes Modell angemessen zu parametrisieren.

Kapitel 5 bietet einen deutschsprachigen Enzyklopädie-Beitrag über die neurokognitive Modellierung (Kapitel 5.1; Jacobs & Hofmann, 2013). Ebenso findet sich in dem Buchkapitel über die Leseforschung mit IAMs eine exemplarische Simulation, die ausführt, wie man die Herausforderung angehen könnte, den Vorschlag dynamischer Aktivierungsausbreitung von

Collins und Loftus (1975) in einem vollständig symbolisch repräsentierenden Modell quantitativ umzusetzen (Abb. 2, Radach & Hofmann, 2017; Kapitel 5.2).

In Kapitel 6.1 simulieren wir ein dynamisches *spreading* assoziativer Energie zwischen Prime- und Zielwort. Um eine realitätsnahe Simulation jeweils zweier Wörter zu ermöglichen, generieren wir ein vollständiges Lexikon für jedes Prime-Zielwort-Paar, das jeweils alle assoziierten Wörter dieser Stimuli enthält. Ein dynamisches AROM erklärt die in Roelke, Franke et al. (2018) beschriebenen antworterleichternden Effekte einer langen SOA dadurch, dass mehr Zeit für assoziative Aktivierungsausbreitung zwischen Prime- und Zielwort-Präsentation zur Verfügung steht. Dies führt zu höheren assoziativen Aktivierungen, bevor das Zielwort präsentiert wird, so dass die Entscheidungskriterien für eine Ja-Antwort schneller erreicht werden können. Ebenso lassen sich Effekte der AS und der Anzahl gemeinsamer Assoziierter simulieren (vgl. Hofmann et al., 2011). Wenn zwischen dem Prime „ECKE" und dem Zielwort „WILLE" weder direkte noch gemeinsame Assoziationen vorliegen, dann resultiert daraus eine lang andauernde semantischen Kompetition (Kapitel 6.1), die vorher bereits verbaltheoretisch postuliert wurde (Thompson-Schill et al., 1997). Diese Simulation erklärt also die IFG-Befunde aus Kapitel 4.4 (vgl. Kapitel 2). In den oszillierenden, semantischen Aktivierungsfunktionen in Abbildung 6A lässt sich die semantische Inhibition und Kompetition zwischen den assoziierten Wörtern von Prime- und Zielwort meines Wissens zum ersten Mal direkt beobachten. Die semantische Kompetition wirkt auf die orthographische Ebene und erklärt damit die Verzögerung der auf Basis dieser Ebene gefällten lexikalischen Entscheidungen (vgl. Grainger & Jacobs, 1996). Mit einem AROM, das um eine episodische Ebene erweitert wurde, ließe sich die Hypothese testen, dass die langanhaltende kompetitive Energie in der semantischen Ebene zum Ausbilden neuer Assoziationen führt – eine CLS-Perspektive, welche die Funktion des DG abbilden sollte (Kumaran & McClelland, 2012; vgl. Kapitel 6.2).

Wenn jedoch eine direkte Assoziation oder gemeinsame Assoziierte zwischen Prime- und Zielwort bestehen (Abb. 6B und 6C), wirken diese semantischen Aktivierungen auch auf die orthographische Ebene, wo wesentlich steiler ansteigende Aktivierungsfunktionen in schnelleren lexikalischen Entscheidungen resultieren (vgl. Abb. 6A).

Kapitel 6.2 basiert auf einem Kommentar zur Quartett-Theorie menschlicher Emotionen und soll insbesondere die Rolle des Hippocampus bei affektiven Prozessen genauer ausführen (Hofmann & Kuchinke, 2015). Die Rolle des Hippocampus beim Verarbeiten von emotionalen Informationen lässt sich dadurch erklären, dass emotionale Konnotation von Wörtern in den

semantischen Strukturen des LZG enthalten sind. Von dort aus fließen diese Informationen in den Hippocampus.

Affektive Informationen und deren Relation zum semantischen LZG werden in verschiedenen Teilen dieser Habilitation behandelt. Hier sollen einige zusammenfassende Schlussfolgerungen daraus gezogen werden: In Hofmann und Jacobs (2014) können wir eine Korrelation der positiven Valenz mit der Anzahl der Assoziationen zu den anderen Wörtern der *Berlin affective word list* aufzeigen (vgl. Kapitel 2; Võ et al., 2009). Ebenso finden wir, dass die positive Valenz keinen Effekt im episodischen Gedächtnis auslöst, der über die Anzahl Assoziierter im Stimulusmaterial hinausgehen würde. Der Effekte negativer Valenz hingegen bleibt neben der Anzahl assoziierter Wörter im Reizmaterial bestehen. So zeigt sich, dass der Effekt positiver, aber nicht negativer Valenz beim Wiedererkennen durch die semantische Kohäsion im Reizmaterial erklärt werden kann (vgl. z. B. Koch, Alves, Krüger, & Unkelbach, 2016). In den geplanten Analysen von Kapitel 4.1 kann die Interaktion der Valenzen zweier Wörter neben der AS einen reproduzierbaren Varianzanteil an den Assoziations-*ratings* aufklären – ein weiterer Beleg für den Zusammenhang von emotionaler Valenz und assoziativen Verknüpfungen zwischen Wörtern. Wenn man für die explorativen Analysen ein *Skip-gram*-Modell in die *linear mixed effects models* als Prädiktor hinzunimmt, dann löst die emotionale Valenz keine reproduzierbaren Effekte mehr aus. Die emotionale Valenz kann als eine Grunddimension des semantischen Raumes betrachtet werden (vgl. Jacobs et al., 2015; Osgood et al., 1957). Deshalb lässt sich die erklärte Varianz der emotionalen Valenz auch durch dieses RNN absorbieren (Hollis et al., 2017; vgl. Westbury et al., 2015). Während unsere symbolische AROM-basierte Definition zumindest die Effekte negativer Valenz noch nicht enthält (Abschnitt 5, Hofmann & Jacobs, 2014), scheint das *Skip-gram*-Modell auch die Effekte negativer Valenz zu enthalten (Hofmann et al., 2018). Dieser Unterschied ließe sich aber auch auf verschiedene kognitive Prozesse beim Wiedererkennen oder Beurteilen der Assoziationsstärke zurückführen. Weil sich relativ abstrakte Wörter häufig auf sehr konkrete, innere Emotionen beziehen (Kousta et al., 2011; Vigliocco et al., 2014), können solche Sprachmodelle ebenso Varianz an *Imageability*-Urteilen, also „Konkretheit“ in der visuellen Domäne, erklären (Hofmann et al., 2018; Westbury et al., 2013).

Solche semantischen Informationen fließen über den entorhinalen Cortex in den Hippocampus (Rolls, 2007). Dort werden neue, rein episodische Wissenskombinationen im DG detektiert, wie sich an Hand einer gemeinsam mit Ralph Radach betreuten Master-Arbeit zeigen ließ: Wenn die Versuchspersonen Wortpaare ohne direkte und gemeinsame Assoziationen lernten,

dann lösten diese Wörter die stärksten Aktivierungen in einer Region aus, die sich dem DG zuordnen ließ (Klein, 2018). Direkte Assoziationen wirken bereits auf einer frühen Stufe der Verarbeitung im Hippocampus. Wenn wir die in ihn einfließenden Afferenzen weiterverfolgen, ist ein weiteres Areal die CA3-Region (Yassa & Stark, 2011). Dort fand Carsten Klein (2018) einen Interaktionseffekt direkter und gemeinsamer Assoziierter. Es konnte also auf Basis dieser computerlinguistischen Definitionen gezeigt werden, dass syntagmatische und paradigmatische (De Saussure, 1959) sowie assoziative und semantische Relationen (Lucas, 2000; Roelke, Franke et al., 2016) jeweils eine physische Entsprechung im Hippocampus haben. Die hierarchische Verarbeitung von zuerst direkten Assoziationen im DG, bevor die direkten Assoziationen mit gemeinsamen Assoziationen in CA3 interagieren, bietet ebenso eine physische Grundlage dafür, diese jeweils als Relationen erster und höherer Ordnung zu bezeichnen (z. B. Landauer & Dumais, 1997). Auch für die in den Hippocampus eingehenden Afferenzen gilt „syntagmatic-first“ (vgl. Friederici, 2002).

Wenn wir uns an die vier Assoziationsgesetze aus Kapitel 1 zurückerinnern und sie mit den Untersuchungen in Kapitel 4 vergleichen, lässt sich festhalten, dass wir komputationale Mechanismen aufgezeigt haben, die beschreiben, wie sich die Gesetze der Kontiguität, Frequenz und Ähnlichkeit in algorithmischen Modellen abbilden lassen. Nur das Gesetz des Kontrastes kann noch nicht überzeugend in symbolisch repräsentierenden Algorithmen umgesetzt werden. Um das Entdecken distinkter semantischer Eigenschaften im DG und damit die Mustertrennung abzubilden (Yassa & Stark, 2011), werden wir in zukünftigen Simulationen XOR-Funktionen verwenden, welche die Mustergröße halbieren und insbesondere distinkte semantische Eigenschaften detektieren (vgl. z. B. Elman, 1990). Dies geschieht durch das Entdecken von semantischen Eigenschaften, die nur zu einem der beiden Stimulus-Wörter eine Assoziation aufweisen. So werden wir den Prozess der Mustertrennung also symbolisch simulieren und das AROM um eine episodische Ebene erweitern. Mit einem Mechanismus der Ähnlichkeit durch die gemeinsamen Assoziierten im AROM und einer Anzeige der nicht gemeinsamen Eigenschaften könnten wir testen, ob sich die meisten taxonomisch-hierarchisch definierbaren Wortrelationen durch ein solches Modell abbilden lassen (Collins & Loftus, 1975; Collins & Quillian, 1969). Eine Übersicht über eine Reihe von *model-to-data connections* hierzu finden sich in Klix (1992, p. 236): Wenn sich assoziierte Wörter als semantische Eigenschaften definieren lassen (vgl. Stuellein et al., 2016), dann sollten Unterbegriffe im Muster der Oberbegriffe größtenteils enthalten sein; Nebenordnungsbegriffe sollten teilweise überlappen; und eine deutlich höhere semantische Überlappung sollte sich bei relativ synonym verwendbaren Begriffen finden (Klix, 1992, p. 236). Darüber hinaus sollte ein auf Kumaran

und McClelland (2012) basierendes Modell des DG auch dazu in der Lage sein, Inferenzen zu beschreiben: Wenn man zum Beispiel lernt, dass ein „Beo“ ein *Vogel* ist und weiß, dass ein „Vogel“ auch *fliegen* kann, dann kann man schlussfolgern, dass ein „Beo“ auch *fliegen* kann. Durch die *conjunction units* sollte man den Prozess der Generalisierung symbolisch abbilden können (vgl. Rumelhart & Todd, 1993, zitiert aus McClelland & Rogers, 2003, p. 314). Mit vollständig symbolisch repräsentierenden Algorithmen ließe sich beispielsweise die Forderung der europäischen Kommission nach transparenten und verständlichen AI-Algorithmen erfüllen. Sie fordert, dass sich Algorithmen der Einsichtnahme durch den Menschen nicht entziehen sollen (*European commission's high-level expert group on artificial intelligence*, 2018, p.14[27]) – eine Forderung, die sich für *hidden units* nur indirekt über die Wirkung auf die Input- und Output-Einheiten erfüllen lässt.

Kapitel 6.3 eröffnet die Perspektive, dass die in Kapitel 4.2 präsentierten Algorithmen Lückentextergänzungsaufgaben lösen können, wie sie ähnlich auch in einen Intelligenztest abgefragt werden (z. B. Liepmann et al., 2007). Somit können wir solchen, häufig auch als AI bezeichneten Algorithmen auch aus psychologischer Perspektive ein gewisses Maß an „*Intelligenz, wie die Tests sie messen*“ zuschreiben (vgl. Titel von Boring, 1923, zitiert aus Amelang & Bartussek, 1997; LeCun et al., 2015). Durch diese Algorithmen wird es möglich, hinreichende Bedingungen zu spezifizieren, die zu einem solchen intelligenten Verhalten führen. Mit den bedingten Wahrscheinlichkeiten des *N-gram*-Modells, die einen hohen Speicherbedarf benötigen, entspricht dieser Algorithmus einer primär gedächtnisbasierten, kristallinen Strategie, die Lückentextergänzungsaufgabe zu lösen. Das SRN hingegen benötigt weniger Speicher, aber dafür mehr Rechenkapazitäten, die darauf verwendet werden, das Wissen in den Verknüpfungen zu wenigen *hidden units* abzubilden – eine Strategie, die man mit der besseren Fähigkeit zu generalisieren näher an den Begriff der fluiden Intelligenz rücken könnte (Klauer et al., 2002). Neben dieser theoretischen Betrachtung des Begriffes „Intelligenz“ haben wir in Kapitel 6.3 auch mögliche Anwendungen algorithmisch konkreter Modelle für die psychologische Praxis skizziert: Es wäre zum Beispiel ein spezifisch für den Interessensbereich eines ADHD-Kindes trainierter Intelligenztest denkbar, der wahrscheinliche Inferenzen dieser Individuen errechnet und Lückentextergänzungsaufgaben spezifisch für diese generiert. So hoffen wir also in der Zukunft einen Beitrag dafür leisten zu können,

[27] „Technological transparency implies that AI systems be auditable,14 comprehensible and intelligible by human beings at varying levels of comprehension and expertise.“

Intelligenztests von Annahmen gleichförmigen Wissens für alle Individuen zu befreien (Catell, 1943)

Zusammengefasst konnte ich zeigen, dass sich vollständig symbolisch repräsentierende Algorithmen sehr konkret mit den vier Assoziationsgesetzen in Zusammenhang bringen lassen. Solche semantischen Technologien sehe ich als einen wichtigen Schritt, um der Forderung nach transparenten Semantik-Modellen gerecht zu werden, bei denen alle Prozesse transparent nachvollzogen werden können (*European commission's high-level expert group on artificial intelligence*, 2018). Dennoch existieren immer noch Daten, die sich durch subsymbolisch repräsentierende Ansätze besser erklären lassen (z. B. Hofmann et al., 2018). Dies liegt meiner Meinung auch daran, dass die *hidden units* in diesen Modellen Abstraktionsprozesse, Generalisierung und Inferenzen derzeit einfacher abbilden können. Insgesamt bieten algorithmische Modelle des semantischen und episodischen Gedächtnisses die Möglichkeit, den Zirkelschluss klassischer operationaler Definitionen zu umgehen, die abhängigen Variablen in der einen Studie zu unabhängigen Variablen in einer zweiten Studie zu machen. Denn durch die Annahmen eines Korpus, das die Erfahrungsgrundlage eines Menschen abbildet, eines Algorithmus, der daraus eine semantische LZG-Struktur konsolidiert, und einer Phase, in welcher der Computer dieselben Aufgaben löst wie der Mensch, kann die psychologische Theoriebildung von diesen hochspannenden Entwicklungen im Bereich AI profitieren.

Literatur

Amelang, M., & Bartussek, D. (1996). *Differentielle Psychologie und Persönlichkeitsforschung*. Stuttgart: Kohlhammer.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463–498.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, *37*(1), 94–117.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*(3), 364–390.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Schmidt, A. M., … Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 449–454.

Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological Review*, *124*(1), 1–20.

Biemann, C., Remus, S., & **Hofmann**, M. J. (2015). Predicting word "predictability" in cloze completion, electroencephalographic and eye movement data. In B. Sharp, W. Lubaszewski, & R. Delmonte (Eds.), *Natural Language Processing and Cognitive Science* (pp. 1–10). Venice, Italy: Libreria Editrice Cafoscarina.

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.

Binder, J. R., McKiernan, K. A., Parsons, M. E., Westbury, C. F., Possing, E. T., Kaufman, J. N., & Buchanan, L. (2003). Neural correlates of lexical access during visual word recognition. *Journal of Cognitive Neuroscience*, *15*(3), 372–393.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Bordag, S. (2007). *Elements of Knowledge-free and Unsupervised Lexical Acquisition* (Doktorarbeit). Universität Leipzig, Leipzig.

Boring, E. G. (1923). Intelligence as the tests test it. *New Republic*, *36*, 35–37.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.

Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. In *Proceedings of GSCL* (pp. 31–40).

Braun, M., Hutzler, F., Ziegler, J. C., Dambacher, M., & Jacobs, A. M. (2009). Pseudohomophone effects provide evidence of early lexico-phonological processing in visual word recognition. *Human Brain Mapping*, *30*(7), 1977–1989.

Braun, M., Jacobs, A. M., Hahne, A., Ricker, B., **Hofmann**, M., & Hutzler, F. (2006). Model-generated lexical activity predicts graded ERP amplitudes in lexical decision. *Brain Research*, *1073–1074*, 431–439.

Brickenkamp, R. (1990). *Die generelle Interessens-Skala (GIS)*. Göttingen: Hogrefe.

Broadbent, D. E. (1967). Word-Frequency Effect and Response Bias. *Psychological Review*, *74*(1), 1–15.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). A Review of Recent Developments and Implications for the Choice of Frequency Estimates in German. *Experimental Psychology*, *58*, 412–424.

Catell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*(3), 153–193.

Cattell, J. M. (1885). Ueber die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben. *Philosophische Studien*, *2*, 635–650.

Chomsky, N. (1959). Reviews: Verbal behavior by B. F. Skinner. *Language*, *35*(1), 26–58.

Chomsky, N. (2002). *Syntactic structures*. New York: Mouton de Gruyter.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.-A., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, *123*, 291–307.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*, 240–247.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*(4), 589–608.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256.

Cowan, N. (2000). The magical number 4 in short-term memory – a reconsideration. *Behavioral and Brain Sciences*, *24*(4), 87–114.

Cox, G. E., & Shiffrin, R. M. (2017). A dynamic approach to recognition memory. *Psychological Review*, *124*(6), 795–860.

Dambacher, M. (2010). *Bottom-up and top-down processes in reading: Influences of frequency and predictability on event-related potentials and eye movements*. Potsdam: Universitätsverlag Potsdam.

Dambacher, M., & Kliegl, R. (2006). Synchronizing timelines: Relations between fixation durations and N400 amplitudes during sentence reading. *Brain Research*, *1155*, 147–162.

Dambacher, M., Dimigen, O., Braun, M., Wille, K., Jacobs, A. M., & Kliegl, R. (2012). Stimulus onset asynchrony and the timeline of word recognition: Event-related potentials during sentence reading. *Neuropsychologia*, *50*(8), 1852–1870.

Dambacher, M., Kliegl, R., **Hofmann**, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, *1084*(1), 89–103.

Dambacher, M., Rolfs, M., Göllner, K., Kliegl, R., & Jacobs, A. M. (2009). Event-related potentials reveal rapid verification of predicted visual input. *PLoS ONE*, *4*(3).

Davis, M. (1992). The role of the amygdala in rear and anxiety. *Annual Review of Neuroscience*, *15*(1), 353–375.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407.

Devlin, J. T., Jamison, H. L., Gonnerman, L. M., & Matthews, P. M. (2006). The role of the posterior fusiform gyrus in reading. *Journal of Cognitive Neuroscience*, *18*(6), 911–922.

Dimigen, O., Kliegl, R., & Sommer, W. (2012). Trans-saccadic parafoveal preview benefits in fluent reading: A study with fixation-related brain potentials. *NeuroImage*, *62*, 381–393.

Dufau, S., Grainger, J., & Ziegler, J. C. (2012). How to say "no" to a nonword: A leaky competing accumulator model of lexical decision. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(4), 1117–1128.

Duncan, K. J., Pattamadilok, C., & Devlin, J. T. (2010). Investigating occipito-temporal contributions to reading with TMS. *Journal of Cognitive Neuroscience*, *22*(4), 739–750.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, *19*, 61–74.

Ebbinghaus, H. (1885). *Über das Gedächtnis*. Leipzig: Duncker & Humblot.

Elman, J. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Eriksson, P. S., Perfilieva, E., Björk-Eriksson, T., Alborn, A.-M., Nordborg, C., Peterson, D., & Gage, D. (1998). Neurogenesis in the adult human hippocampus. *Nature*, *4*(11), 1313–1317.

European commission's high-level expert group on artificial intelligence. (2018). Draft: Ethics guidelines for trustworthy AI. Gefunden am 8. Februar 2019 unter https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=57112

Evert, S. (2005). *The Statistics of Word Cooccurrences Word Pairs and Collocations* (Doktorarbeit). Universität Stuttgart, Stuttgart.

Ferrand, L., & New, B. (2003). Semantic and associative priming in the mental lexicon. In P. Bonin (Ed.), *Mental lexicon: "Some words to talk about words"* (pp. 25–43). New York: Nova science publishers.

Feynman, R. (1968). Richard Feynman on how we would look for a new law (the key to science). Gefunden am 12. Juni 2018, unter http://amiquote.tumblr.com/post/4463599197/richard-feynman-on-how-we-would-look-for-a-new-law

Fiebach, C. J., Friederici, A. D., & Cramon, D. Y. (1999). fMRI evidence for dual routes to the mental lexicon in visual word recognition. *Journal of Cognitive Neuroscience*, 11–23.

Forgács, B., Bohrn, I., Baudewig, J., **Hofmann**, M. J., Pléh, C., & Jacobs, A. M. (2012). Neural correlates of combinatorial semantic processing of literal and figurative noun noun compound words. *NeuroImage*, *63*, 1432–1442.

Fox, P. T., & Lancaster, J. L. (2002). Mapping context and content: The BrainMap model. *Nature Reviews Neuroscience*, *3*(4), 319–321.

Frank, S. L., & Willems, R. M. (2017). Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. *Language, Cognition and Neuroscience*, *32*(9), 1192–1203.

Franke, N., Roelke, A., Radach, R. R., & **Hofmann**, M. J. (2017). After braking comes hasting: Reversed effects of indirect associations in 2nd and 4th graders. In *Proceedings of the Cognitive Science Society* (pp. 2025–2030). London, UK.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493–501.

Gamallo, P. (2018). CitiusNLP at SemEval-2018 Task 10: The use of transparent distributional models and salient contexts to discriminate word attributes. In *Proceedings of The 12th International Workshop on Semantic Evaluation* (pp. 953–957). New Orleans, Lousiana, USA: Association for Computational Linguistics.

Giora, R., Kronrod, A., Elnatan, I., Shuval, N., & Zur, A. (2004). Weapons of mass distraction: Optimal innovation and pleasure ratings. *Metaphor & Symbol*, *19*(2), 115–141.

Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 500–513.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychological Review*, *103*(3), 518–565.

Grainger, J., & Jacobs, A. M. (1998). Localist connectionist approaches to human cognition. Mahwah, USA: Lawrence Erlbaum.

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighbourhood frequency effect. *Perception and Psychophysics*, *45*(3), 189–195.

Green, D. M., & Swets, J. A. (1966). *Signal detection and psychophysics*. New York, USA: Wiley.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211–244.

Grossberg, S. (1978). A theory of visual coding, memory, and development. In E. Laurens, H. F. Leeuwenberg, & J. M. Buffart (Eds.), *Formal Theories of Visual Perception* (pp. 7–26). New York, USA.

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun – An R package for computations based on latent semantic analysis. *Behavior Research Methods*, *47*(4), 930–944.

Guthrie, E. R. (1930). Conditioning as a prinicple of learning. *Psychological Review*, *37*(5), 412–428.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*(3), 662–720.

Harris, Z. S. (1963). *Structural Linguistics*. Chicago, USA: The university of Chicago press.

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, *115*(5), 1090–1103.

Hebb, D. O. (1949). *The organisation of behavior: A neuropsychological theory*. New York: John Wiley & Sons, Inc.

Hempel, C. G., & Oppenheim, P. (1948). Studies on the logic of explanation. *Philosophy of Science*, *15*(2), 135–175.

Hofmann, M. J., & Jacobs, A. M. (2014). Interactive activation and competition models and semantic context: From behavioral to brain data. *Neuroscience and Biobehavioral Reviews*, *46*, 85–104.

Hofmann, M. J., & Kuchinke, L. (2015). "Anything is good that stimulates thought" in the hippocampus: Comment on "The quartet theory of human emotions: An integrative and neurofunctional model" by S. Koelsch et al. *Physics of Life Reviews*, *13*, 58–60.

Hofmann, M. J., Biemann, C., & Remus, S. (2017). Benchmarking n-grams, topic models and recurrent neural networks by cloze completions, EEGs and eye movements. In B.

Sharp, F. Sedes, & W. Lubaszewsk (Eds.), *Cognitive Approach to Natural Language Processing* (pp. 197–215). London, UK: ISTE Press Ltd, Elsevier.

Hofmann, M. J., Biemann, C., Westbury, C. F., Murusidze, M., Conrad, M., & Jacobs, A. M. (2018). Simple co-occurrence statistics reproducibly predict association ratings. *Cognitive Science*, *42*, 2287–2312.

Hofmann, M. J., Dambacher, M., Jacobs, A. M., Kliegl, R., Radach, R., Kuchinke, L., … Herrmann, M. J. (2014). Occipital and orbitofrontal hemodynamics during naturally paced reading: An fNIRS study. *NeuroImage*, *94*, 193–202.

Hofmann, M. J., Herrmann, M. J., Dan, I., Obrig, H., Conrad, M., Kuchinke, L., … Fallgatter, A. J. (2008). Differential activation of frontal and parietal regions during visual word recognition: An optical topography study. *NeuroImage*, *40*, 1340–1349.

Hofmann, M. J., Kuchinke, L., Biemann, C., Tamm, S., & Jacobs, A. M. (2011). Remembering words in context as predicted by an associative read-out model. *Frontiers in Psychology*, *2*(252), 1–11.

Hofmann, M. J., Kuchinke, L., Tamm, S., Võ, M.-L., & Jacobs, A. M. (2009). Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(4), 389–397.

Hofmann, M. J., Tamm, S., Braun, M. M., Dambacher, M., Hahne, A., & Jacobs, A. M. (2008). Conflict monitoring engages the mediofrontal cortex during nonword processing. *Neuroreport*, *19*(1), 25–29.

Holcomb, P. J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, *14*(6), 938–950.

Hollis, G., Westbury, C., & Lefsrud, L. (2017). Extrapolating human judgments from skip-gram vector representations of word meaning. *Quarterly Journal of Experimental Psychology*, *70*(8), 1603–1619.

Huber, D. E., Curran, T., O'Reilly, R. C., & Woroch, B. (2008). The dynamics of integration and separation: ERP, MEG, and neural network studies of immediate repetition effects. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1389–1416.

Huey, E. B. (1908). *The psychology and pedagogy of reading*. New York, USA: The Macmillan Company.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap ? A micro analytic review. *Psychonomic Bulletin & Review*, *10*(4), 785–813.

Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C. S., … Buchanan, E. (2013). The semantic priming project. *Behavior Research Methods*, *45*(4), 1099–1114.

Hutzler, F., Braun, M., Võ, M. L. H., Engl, V., **Hofmann**, M., Dambacher, M., … Jacobs, A. M. (2007). Welcome to the real world: Validating fixation-related brain potentials for ecologically valid settings. *Brain Research*, *1172*(1), 124–129.

Hyde, H. M. (1962). *The trials of Oscar Wilde*. New York, USA: Dover Publications Inc.

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception* (pp. 29–53). Oxford, England: Elsevier Science.

Inhoff, A. W., Radach, R., Eiter, B. M., & Juhasz, B. (2003). Distinct subsystems for the parafoveal processing of spatial and linguistic information during eye fixations in reading. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *56*(5), 803–827.

Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1311–1334.

Jacobs, A. M., & **Hofmann**, M. J. (2013). Neurokognitive Modellierung. In E. Schröger & S. Koelsch (Eds.), *Enzyklopädie der Psychologie. Affektive und kognitive Neurowissenschaft* (pp. 431–447). Göttingen: Hogrefe.

Jacobs, A. M., Graf, R., & Kinder, A. (2003). Receiver operating characteristics in the lexical decision task: Evidence for a simple signal-detection process simulated by the multiple read-out model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 481–488.

Jacobs, A. M., **Hofmann**, M. J., & Kinder, A. (2016). On elementary affective decisions: To like or not to like, that is the question. *Frontiers in Psychology*, *7*(1836), 1–17.

Jacobs, A. M., Võ, M. L., Briesemeister, B. B., Conrad, M., **Hofmann**, M. J., & Kuchinke, L. (2015). 10 years of BAWLing into affective and aesthetic processes in reading: what are the echoes? *Froniers in Psychology*, *6*(714), 1–15.

Jordan, M. I. (1986). Serial order: A parallel distributed processing approach (Tech. Rep. No. 8604). San Diego, USA: University of California, Institute for Cognitive Science.

Jung-Merker, L., & Rüf, E. (2011). *C. G. Jung: Experimentelle Untersuchungen*. Ostfildern: Patmos Verlag der Schwabenverlag AG.

Jurgens, D. A., Turney, P. D., Mohammad, S. M., & Holyoak, K. J. (2012). Semeval-2012 task 2: Measuring degrees of relational similarity. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics* (pp. 356–364). Montreal, Canada.

Kersting, M., Althoff, K., & Jäger, A. O. (2008). *WIT-2: Wilde-Intelligenz-Test 2*. Göttingen: Hogrefe.

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, *48*(7), 805–825.

Kimball, D. R., Smith, T. A., & Kahana, M. J. (2007). The fSAM model of false recall. *Psychological Review*, *114*(4), 954–993.

Klauer, K. J., Willmes, K., & Phye, G. D. (2002). Inducing inductive reasoning: Does it transfer to fluid intelligence? *Contemporary Educational Psychology*, *27*(1), 1–25.

Klein, C. (2018). *Hippocampale Plastizität bei der Konsolidierung neuartiger Gedächtnisengramme: Eine fMRT – Studie* (Masterarbeit). Bergische Universität Wuppertal, Wuppertal.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1–2), 262–284.

Klix, F. (1992). *Die Natur des Verstandes*. Göttingen: Hogrefe.

Kneser, R., & Ney, H. (1995). Improved backing-off for m-gram language modeling. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 181–184). Detroit, MI, USA: IEEE.

Koch, A., Alves, H., Krüger, T., & Unkelbach, C. (2016). A general valence asymmetry in similarity: Good is more alike than bad. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1171–1192.

Koelsch, S., Jacobs, A. M., Menninghaus, W., Liebal, K., Klann-Delius, G., von Scheve, C., & Gebauer, G. (2015). The quartet theory of human emotions: An integrative and neurofunctional model. *Physics of Life Reviews*, *13*, 1–27.

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34.

Kronbichler, M., Hutzler, F., Wimmer, H., Mair, A., Staffen, W., & Ladurner, G. (2004). The visual word form area and the frequency with which words are encountered: evidence from a parametric fMRI study. *NeuroImage*, *21*(3), 946–953.

Kuchinke, L., Fritzemeier, S., **Hofmann**, M. J., & Jacobs, A. M. (2013). Neural correlates of episodic memory: Associative memory and confidence drive hippocampus activations. *Behavioural Brain Research*, *254*, 92–101.

Kuchinke, L., **Hofmann**, M. J., Jacobs, A. M., Frühholz, S., Tamm, S., & Herrmann, M. (2011). Human striatal activation during adjustment of the response criterion in visual word recognition. *NeuroImage*, *54*, 2412–2417.

Kuhlmann, M., **Hofmann**, M. J., & Jacobs, A. M. (2017). If you don't have valence, ask your neighbor: Evaluation of neutral words as a function of affective semantic associates. *Frontiers in Psychology*, *8*(343), 1–7.

Kuhlmann, M., **Hofmann**, M. J., Briesemeister, B., & Jacobs, A. M. (2016). Mixing positive and negative valence: Affective-semantic integration of bivalent words. *Scientific Reports*, *6*(30718), 1–7.

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review*, *119*(3), 573–616.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Larsson, J., & Heeger, D. J. (2006). Two retinotopic visual areas in human lateral occipital cortex. *Journal of Neuroscience*, *26*(51), 13128–13142.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews. Neuroscience*, *9*(12), 920–933.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, *3*, 211–225.

Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *I-S-T 2000R: Intelligenz-Struktur-Test 2000R*. Göttingen: Hogrefe.

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, *7*(4), 618–630.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, *88*, 22–60.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, Massachusetts: The MIT Press.

Maratos, E. J., Allan, K., & Rugg, M. D. (2000). Recognition memory for emotionally negative and neutral words: an ERP study. *Neuropsychologia*, *38*(11), 1452–1465.

Marr, D. (1970). A theory for cerebral neocortex. *Proceedings of the Royal Society B: Biological Sciences*, *176*(1043), 161–234.

Marr, D. (1971). Simple Memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *262*(841), 23–81.

Marsman, J. B. C., Renken, R., Velichkovsky, B. M., Hooymans, J. M. M., & Cornelissen, F. W. (2012). Fixation based event-related fMRI analysis: Using eye fixations as

events in functional magnetic resonance imaging to reveal cortical processing during the free exploration of visual images. *Human Brain Mapping*, *33*(2), 307–318.

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General*, *142*(4), 1190–1210.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, *4*(4), 310–322.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*(5), 375–407.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457.

McInerney, R. J., & Kerns, K. A. (2003). Time reproduction in children with ADHD: motivation matters. *Child Neuropsychology*, *9*(2), 91–108.

McKeon, R. (1941). *The basic works of Aristotle*. New York: Random House.

McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *18*(6), 1155–1172.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. New York: Taylor & Francis Group.

Mecklinger, A. (2010). The control of long-term memory: Brain systems and cognitive processes. *Neuroscience and Biobehavioral Reviews*, *34*(7), 1055–1065.

Mikolov, T. (2012). *Statistical language models based on neural networks* (Doktorarbeit). Brno university of technology, Brno, Czech Republic.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. Gefunden unter https://arxiv.org/abs/1301.3781.

Mikolov, T., Le, Q. V., & Sutskever, I. (2013). Exploiting similarities among languages for machine translation. Gefunden unter https://arxiv.org/abs/1309.4168.

Mikolov, T., Yih, W., & Zweig, G. (2013). Linguistic regularities in continuous space word representations. In *Proceedings of NAACL-HLT* (pp. 746–751). Atlanta, GA, USA.

Miller, G. A. (1994). The magical number seven, plus or minus two: Some limits on our capacity for processing, *101*(2), 343–352.

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141–144.

Minsky, M. L. (1961). Steps toward artificial intelligence. *Proceedings of the IRE*, *49*(1), 8–30.

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). Semantic significance: A new measure of feature salience. *Memory and Cognition*, *42*(3), 355–369.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226–254.

Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: A review of research and theory. *Psychological Bulletin*, *84*(1), 93–116.

New., B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English lexicon project. *Psychonomic Bulletin and Review*, *13*(1), 45–52.

Nobre, A. C., Coull, J. T., Frith, C. D., & Mesulam, M. M. (1999). Orbitofrontal cortex is activated during breaches of expectation in tasks of visual attention. *Nature Neuroscience*, *2*(1), 11–12.

O'Reilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition. *Trends in Cognitive Sciences*, *2*(11), 455–462.

Oganian, Y., Froehlich, E., Schlickeiser, U., **Hofmann**, M. J., Heekeren, H. R., & Jacobs, A. M. (2016). Slower perception followed by faster lexical decision in longer words: A diffusion model analysis. *Frontiers in Psychology*, *6*(1958), 1–12.

Olson, M. H., & Hergenhahn, B. R. (2017). *An introduction to theories of learning*. New York, USA: Routledge.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, Illinois, USA: University of Illinois press.

Page, M. (2000). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences*, *23*(4), 443–467.

Panchenko, A., Ruppert, E., Faralli, S., Ponzetto, S. P., & Biemann, C. (2017). Unsupervised does not mean uninterpretable: The case for word sense induction and disambiguation. In *Proceedings of EACL* (Vol. 1, pp. 86–98). Valencia, Spain: Association for Computational Linguistics.

Pasche, S. (2018). *Individuelle Wortfrequenz als Prädiktor visueller Worterkennung: Eine Tablet-Based Eye Tracking Studie* (Bachelorarbeit). Bergische Univresität Wuppertal, Wuppertal.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews. Neuroscience*, *8*(12), 976–987.

Perea, M., & Carreiras, M. (2003). Sequential effects in the lexical decision task: The role of the item frequency of the previous trial. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *56*(3), 385–401.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*(2), 273–315.

Phan, X.-H., & Nguyen, C.-T. (2007). GibbsLDA++: A C/C++ implementation of latent Dirichlet allocation (LDA). Gefunden am 5.3.2019 unter http://gibbslda.sourceforge.net.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review*, *107*(4), 786–823.

Poeppel, D. (1996). A critical review of PET studies of phonological processing. *Brain & Language*, *55*(3), 317–385.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63.

Popper, K. (1935). *Logik der Forschung*. Berlin: Springer.

Price, C. J. (2000). The anatomy of language: Contributions from functional neuroimaging. *Journal of Anatomy*, *197*(3), 335–359.

Price, C. J., & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, *15*(6), 246–253.

Price, C. J., & Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology*, *22*(3), 262–275.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351–360.

Quasthoff, U., Richter, M., & Biemann, C. (2006). Corpus portal for search in monolingual corpora. In *Proceedings of LREC-06* (pp. 10–13). Genova.

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science*, *12*(5), 410–430.

Quillian, R. (1962). A design for an understanding machine. *Mechanical Translation*, *7*(1), 17–29.

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, *2*(9), 693–705.

Radach, R., & **Hofmann**, M. J. (2016). Graphematische Verarbeitung beim Lesen von Wörtern. In U. Domahs & B. Primus (Eds.), *Laut, Gebärde, Buchstabe (Handbuch Sprachwissen, Band 2)* (pp. 455–473). Berlin: De Gruyter Mouton.

Radach, R., Inhoff, A. W., Glover, L., & Vorstius, C. (2013). Contextual constraint and N + 2 preview effects in reading. *Quarterly Journal of Experimental Psychology*, *66*(3), 619–633.

Radach, R., Inhoff, A., & Heller, D. (2004). Orthographic regularity gradually modulates saccade amplitudes in reading. *European Journal of Cognitive Psychology*, *16*(1–2), 27–51.

Rahman, R. A., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes*, *24*(5), 713–734.

Rapp, R. (2002). The computation of word associations: Comparing syntagmatic and paradigmatic approaches. In *Proceedings of the 19th international conference on Computational Linguistics* (Vol. 1, pp. 1–7). Taipe, Taiwan: Association for Computational Linguistics.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: Comparisons to other models. *The Behavioral and Brain Sciences*, *26*(4), 445–476.

Reilly, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research*, *7*, 34–55.

Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*(2), 245–266.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814.

Roelke, A., Franke, N., Biemann, C., Radach, R., Jacobs, A. M., & **Hofmann**, M. J. (2018). A novel co-occurrence-based approach to predict pure associative and semantic priming. *Psychonomic Bulletin and Review*, *25*(4), 1488–1493.

Roelke, A., Franke, N., Radach, R., Jacobs, A. M., & **Hofmann**, M. J. (2016). *Semantic higher order but not direct associations prime ventral visual stream activation*. Manuskript zur Veröffentlichung eingereicht.

Roelke, A., **Hofmann**, M. J., & Radach, R. (2018). Retinotopic mapping of parafoveal preview during reading? A fixation-related NIRS study. Poster presented at 59th Annual Meeting of the Psychonomic Society. Gefunden unter https://www.researchgate.net/publication/328967301_Retinotopic_mapping_of_parafoveal_preview_during_reading_A_fixation-related_NIRS_study.

Rolls, E. T. (2007). An attractor network in the hippocampus. *Learning & Memory*, *14*(11), 714–731.

Rosenblatt, F. (1958). The Perceptron: A probabilistic model for informationm storage and organization in the brain. *Psychological Review*, *65*(6), 386–408.

Rossell, S. L., Price, C. J., & Nobre, A. C. (2003). The anatomy and time course of semantic priming investigated by fMRI and ERPs. *Neuropsychologia*, *41*(5), 550–564.

Rubenstein, H., & Goodenough, J. B. (1965). Contextual correlates of synonymy. *Communications of the ACM*, *8*(10), 627–633.

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 487–494.

Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representations. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance 14: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 3–30). Cambridge, Massachusetts, USA: The MIT Press.

Saussure, F. de. (1959). *Course in general linguistics*. (C. Bally & A. Sechehaye, Eds.). New York: Philosophical Library.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66.

Schurz, M., Sturm, D., Richlan, F., Kronbichler, M., Ladurner, G., & Wimmer, H. (2010). A dual-route perspective on brain activation in response to visual words: Evidence for a length by lexicality interaction in the visual word form area (VWFA). *NeuroImage*, *49*, 2649–2661.

Schvaneveldt, R., Meyer, D., & Becker, C. (1976). Lexical ambiguity, semantic context, and visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance.*, *2*(2), 243–250.

Seabala [Nutzername]. (2018). Feynman on scientific method. Gefunden am 6. Februar 2019 unter https://www.youtube.com/watch?v=EYPapE-3FRw

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568.

Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time line of word recognition: Evidence from eye movements and event-related potentials. *Neuroreport*, *9*, 2195–2200.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423, 623–656.

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist*, *26*(2), 145–159.

Skinner, B. F. (1948). Verbal behavior: William James lectures. Gefunden am 6. Februar 2019, unter https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwjAi4vu96bgAhWDKVAKHaINDl0QFjAAegQIARAC&url=http%3A%2F

%2Fbfskinner.org%2Fwp-content%2Fuploads%2F2014%2F02%2FWilliam-James-Lectures.pdf&usg=AOvVaw0NkFg4b_gwPUjLwHQ0ODQ

Snyder, J. S. (2019). Recalibrating the relevance of adult neurogenesis. *Trends in Neurosciences*, *42*(3), 164–178.

Sorrells, S. F., Paredes, M. F., Cebrian-Silla, A., Sandoval, K., Qi, D., Kelley, K. W., … Alvarez-Buylla, A. (2018). Human hippocampal neurogenesis drops sharply in children to undetectable levels in adults. *Nature*, *555*(7696), 377–381.

Squire, L. R., & Wixted, J. T. (2011). The cognitive neuroscience of human memory since H. M. *Annual Review of Neuroscience*, *34*, 259–288.

Stuellein, N., Radach, R. R., Jacobs, A. M., & **Hofmann**, M. J. (2016). No one way ticket from orthography to semantics in recognition memory: N400 and P200 effects of associations. *Brain Research*, *1639*, 88–98.

Stumm, S. von, & Plomin, R. (2015). Socioeconomic status and the growth of intelligence from infancy through adolescence. *Intelligence*, *48*, 30–36.

Taylor, W. L. (1953). "Cloze" procedure: A new tool for measuring readability. *Journalism Quarterly*, *30*(4), 415–433.

Thompson-Schill, S. L., & Botvinick, M. M. (2006). Resolving conflict: A response to Martin and Cheng (2006). *Psychonomic Bulletin and Review*, *13*(3), 402–408.

Thompson-Shill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(26), 14792–14797.

Thompson, W. H. (1902). *The work of the digestive glands: Lectures by J. P. Pawlow*. Philadelphia: J. B. Lippincott company.

Thorndyke, A. M. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review*, *2*(4), 1–109.

Trumpp, N. M., Traub, F., & Kiefer, M. (2013). Masked priming of conceptual features reveals differential brain activation during unconscious access to conceptual action and sound information. *PLoS ONE*, *8*(5), 1–10.

Vigliocco, G., Kousta, S. T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: The role of emotion. *Cerebral Cortex*, *24*(7), 1767–1777.

Võ, M. L. H., Conrad, M., Kuchinke, L., Urton, K., **Hofmann**, M. J., & Jacobs, A. M. (2009). The Berlin affective word list reloaded (BAWL-R). *Behavior Research Methods*, *41*(2), 534–538.

Wagenmakers, E. J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, *58*(1), 140–159.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158–177.

Westbury, C. (2016). Pay no attention to that man behind the curtain. *The Mental Lexicon*, *11*(3), 350–374.

Westbury, C. F., Shaoul, C., Hollis, G., Smithson, L., Briesemeister, B. B., **Hofmann**, M. J., & Jacobs, A. M. (2013). Now you see it, now you don't: On emotion, context, and the algorithmic prediction of human imageability judgments. *Frontiers in Psychology*, *4*(991), 1–13.

Westbury, C., Keith, J., Briesemeister, B. B., **Hofmann**, M. J., & Jacobs, A. M. (2015). Avoid violence, rioting, and outrage; approach celebration, delight, and strength: Using large text corpora to compute valence, arousal, and the basic emotions. *The Quarterly Journal of Experimental Psychology*, *68*(8), 1599–1622.

Wettler, M., & Rapp, R. (1989). A connectionist system to simulate lexical decisions in information retrieval. In R. Pfeifer, Z. Schreter, F. Fogelman-Soulié, & L. Steels (Eds.), *Connectionismin perspective* (pp. 463–469). Amsterdam: Elsevier Science.

Wettler, M., & Rapp, R. (1993). Computation of word associations based on the co-occurences of words in large corpora. In *Proceddings of the workshop very large corpora: Academic and industrial perspectives* (pp. 84–93). Columbus, Ohio,USA: Association for Computational Linguistics.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.

Wood, S. N. (2011). Fast stable REML and ML estimation of semiparametric GLMs. *Ournal of the Royal Statistical Society, Series B (Statistical Methodology)*, *73*(1), 3–36.

Yarkoni, T., Speer, N. K., Balota, D. A., McAvoy, M. P., & Zacks, J. M. (2008). Pictures of a thousand words: Investigating the neural mechanisms of reading with extremely rapid event-related fMRI. *NeuroImage*, *42*, 973–987.

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*(4), 931–959.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1341–1354.

Zhila, A., Yih, W., Meek, C., Zweig, G., & Mikolov, T. (2013). Combining heterogeneous models for measuring relational similarity. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1000–1009). Atlanta, GA, USA: Association for Computational Linguistics.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29.

Ziegler, J. C., Jacobs, A. M., & Klüppel, D. (2001). Pseudohomophone effects in lexical decision: Still a challenge for current word recognition models. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 547–559.

Danksagung

Hiermit möchte ich mich herzlich bei allen bedanken, die mich in meinem wissenschaftlichen Wirken von der Entstehung der ersten Schriften dieser Habilitation bis hin zum Abschluss des Verfahrens begleitet haben:
Ralph Radach für den Spaß beim Diskutieren, das viele neue Wissen und die Inspiration in Richtung Praxis weiterzudenken;
Arthur Jacobs für die Überzeugungskraft, neurokognitive Prozesse durch Simulationsmodelle zu verstehen;
Chris Biemann, dass er mir die wunderbare Welt computerlinguistischer Modelle gezeigt hat;
Lars Kuchinke, der mir hilft die psychologischen Mechanismen hinter diesen Modellen zu erkennen;
Andre Rölke und Nicole Franke – ich durfte ihnen die Wissenschaft näherbringen und bin ihr dadurch selbst nähergekommen;
Allen Ko-Autorinnen und Ko-Autoren, die mir so viele verschiedene Perspektiven auf die Wissenschaft eröffnet haben;
Der Deutschen Forschungsgemeinschaft für die finanzielle Unterstützung (DFG-Gz. HO 5139/2-1 und 2-2);
Der Habilitationskommission für das Vortragsthema „Gibt es eine zentrale Exekutive im Gehirn?“, insbesondere auch dem Vorsitzenden, Ralf Schulze, dass dieses Verfahren trotz der widrigen Umstände erfolgreich abgeschlossen werden konnte;
Rasha Abdel-Rahman und Markus Kiefer, dass sie diese umfangreiche Arbeit gelesen, begutachtet, und mich in meinem Weg bestätigt haben;
Meiner Familie, durch die genügend Herzblut in mein Gehirn gepumpt wird – allen voran Gaby für die Perspektive einer fachfremden Leserin.

Wuppertal, den 12.5.2021

…….

(Markus Hofmann)