

# Erstellung und Kontrolle weiterer Stimuli der MIST-Datenbank

Sina Koller, Nadine Müller & Christina Kauschke  
Philipps-Universität Marburg  
Institut für Germanistische Sprachwissenschaft

## 1. Hintergrund: MIST-Datenbank und DFG-Projekt „Durch die rosarote Brille“

Das vorliegende Dokument beschreibt die Erweiterung der Datenbank „Metaphors for Internal State Terms“ (MIST), die im Rahmen des DFG-Projekts „Durch die rosarote Brille“ entstanden ist. Die MIST-Datenbank wurde mit dem Ziel aufgebaut, umfassend kontrollierte Stimuli für die empirische Erforschung des Verständnisses und der Verarbeitung metaphorischer *Internal State Language* bereitzustellen. Für eine Beschreibung der MIST-Datenbank, ihrer Entwicklung und ihres Aufbaus sei auf Müller et al. (2021) und die [Webseite der Datenbank](#) sowie des [Metaphernprojekts der AG Klinische Linguistik der Universität Marburg](#) verwiesen.

## 2. Erweiterung der MIST-Datenbank

Um den Umfang der Datenbank MIST zu erweitern, wurden 168 neue Stimulisets, die auf Metaphern mit sieben zusätzlichen, nicht primär sensorisch basierten Quelldomänen beruhen, generiert und analog zum Vorgehen in Müller et al. (2021) genormt. Das Ergebnis war folglich eine Verdoppelung des Umfangs der bisherigen Datenbank mit vergleichbaren Normierungswerten aller Einträge, die im Rahmen der Masterarbeit Sina Kollers (Koller 2021) erfolgte.

### 2.1 Methodik

#### 2.1.1 Material

Für die Erweiterung der Datenbank MIST wurden die Quelldomänen Körper passiv/Interozeption, Tier, Weg, Richtung/Orientierung, Kraft, Technik und Haushalt gewählt. Während die Stimuli in Müller et al. (2021) sensorisch basieren, lässt sich auch vor allem in den Kategorien Körper passiv/Interozeption, Weg und Orientierung/Ausrichtung klar *embodiment* erkennen, das im Gegensatz zu den Stimuli in Müller et al. (2021) nicht über die Primärsinne kodiert ist. Auch die übrigen Quelldomänen beziehen sich auf Erfahrungen des Körpers, beziehungsweise dessen Aktionen, mit sich selbst und seiner Umwelt auf unterschiedlich abstrakten Niveaus.

Die Stimuli wurden analog zum Vorgehen in Müller et al. (2021) ausgewählt und in Kontext gesetzt. Ebenso wurden analog zu Müller et al. (2021) Ratings durchgeführt, um die Übereinstimmung der metaphorischen Ausdrücke mit ihren jeweiligen wörtlichen Ausdrücken zu bestimmen. An diesen Ratings beteiligten sich 179 Personen, von denen 18 keine gültigen Datensätze produzierten. Die Mittelwerte der Übereinstimmung zwischen Metapher und jeweiliger wörtlicher Entsprechung wurden folglich aus 161 validen Datensätzen berechnet und waren durchweg hoch. Die Werte der Übereinstimmung aller 168 Stimulisets sind in Abb. 1 dargestellt. Beispiele der Stimuli aller neuen sieben Quelldomänen sind in Tab. 1 angeführt.

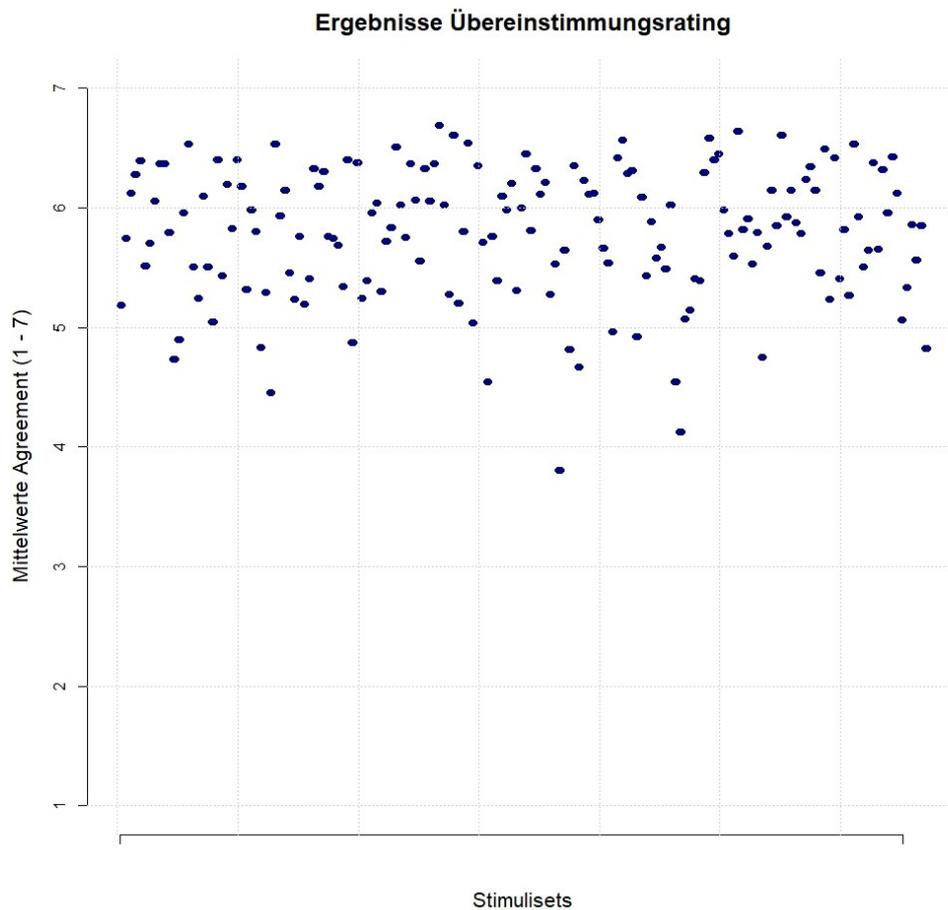


Abb. 1: Mittelwerte der Übereinstimmung zwischen wörtlichen und metaphorischen Stimuli (n = 168).

Tab. 1: Metaphorische und wörtliche Beispiele der sieben Quelldomänen der neu entwickelten Stimuli.

Quelldomäne	Metapher	Wörtliche Entsprechung	Kontext
Haushalt	gebacken bekommen	hinkriegen	Ordnung fiel ihr schwer. Viele Dinge hat sie deshalb nie gebacken bekommen. / hingekriegt.
Körper passiv/ Interozeption	gekränkt	beleidigt	Er war sehr unhöflich. Jetzt ist sie gekränkt. / beleidigt.
Kraft	unbeschwert	froh	Heute hat sie frei. Sie ist ganz unbeschwert. / froh.
Orientierung/ Ausrichtung	Hochgefühl	Euphorie	Das Team hat gewonnen. Es herrscht tagelang ein Hochgefühl. / Euphorie.
Tier	bockig	trotzig	Das Kleinkind quengelt viel. Heute ist es besonders bockig. / trotzig.
Technik	aurasten	sich aufregen	Der Schiedsrichter zeigt die rote Karte. Die Fans rasten aus. / regen sich auf.
Weg	Fortschritt	Verbesserung	Seit Kurzem schreibt sie gute Noten. Sie ist stolz auf ihren Fortschritt. / ihre Verbesserung.

### 2.1.2 Teilnehmer\*innen

Die Werte der Valenz, Familiarität und des Arousals wurden über Fragebögen, die über die Online-Plattform SoSci Survey (Leiner 2019b) erstellt und bearbeitet wurden, erhoben. Der Zeitraum der Befragung belief sich auf 67 Tage im August bis Oktober 2020. Die digitalen Fragebögen wurden nach Vorlage von Müller et al. (2021) erarbeitet.

Um den Arbeitsaufwand für die Teilnehmer\*innen erträglich zu halten und den Zeitrahmen zu begrenzen, wurden pro zu bewertender Variable zwei Fragebögen mit jeweils der Hälfte der zu bewertenden Items erstellt. Um eine gute statistische Aussagekraft der Ergebnisse zu erzielen, wurden mindestens 80 Teilnehmer\*innen pro Fragebogen angestrebt, folglich insgesamt 480. Jede Person wurde über einen der sechs eigentlichen Fragebögen übergeordneten Verteilerbogen gleichverteilt zufällig einem der sechs Fragebögen zugewiesen. Gegen Ende des Befragungszeitraums wurde die Verteilungsgewichtung der Teilnehmer\*innen zugunsten der Fragebögen mit geringeren Abschlussquoten angepasst.

Zur Teilnahme an den Fragebögen wurde primär über den universitären E-Mail-Verteiler der Philipps-Universität Marburg aufgerufen. Auch im Bekannten- und Familienkreis, über gängige Social-Media-Plattformen (Facebook, WhatsApp), in Aushängen und über Survey-Sharing-Plattformen (surveycircle.com, surveyswap.io) wurden Teilnehmer\*innen gewonnen. Unabhängig der Kommunikationskanäle wurden die Einschlusskriterien stets klar beschrieben: Alle Teilnehmer\*innen mussten zwischen 18 und 70 Jahren alt sein und Deutsch als Muttersprache sprechen. Unter allen Teilnehmer\*innen wurden acht 25€-Gutscheine von [wunschgutschein.de](http://wunschgutschein.de) verlost.

Insgesamt beteiligten sich 616 Personen an der Befragung. 102 dieser Teilnehmer\*innen mussten aus verschiedenen Gründen ausgeschlossen werden: 24 Personen gaben an, Deutsch nicht als Muttersprache gelernt zu haben, und 78 Personen beantworteten die Aufmerksamkeitsfragen entweder teilweise oder vollständig fehlerhaft. Es wurde zusätzlich die Qualität relativ zur Bearbeitungsgeschwindigkeit der Fragebögen überprüft, wofür die von SoSci Survey bereitgestellten Qualitätskontrollenkennwerte *relative speed index* (RSI, Leiner 2019a) und „deg time“ (Leiner 2020) verwendet wurden. Durch diese Qualitätskontrolle mussten jedoch keine weiteren Fälle ausgeschlossen werden. Folglich belief sich die Anzahl valider Teilnehmer\*innen auf 514 Personen. Insgesamt berechneten sich die Werte für die Variable Arousal letztendlich aus 173 Datensätzen, die Werte für die Variable Familiarität aus 175 Datensätzen und die Werte für die Variable Valenz aus 166 Datensätzen.

Die Teilnehmer\*innen waren in der Mehrheit weiblich (367 weiblich; 141 männlich; 6 anderes). Da die Teilnehmer\*innen ihr Alter nicht als Freitext angaben, sondern das Alter in Altersspannen erhoben wurde, lässt sich kein exaktes Durchschnittsalter angeben. Der Median der Altersverteilung liegt im Bereich 20–25 Jahre; das mittlere Alter lässt sich auf 26,7 bis 31,5 Jahre eingrenzen. Zudem zeichneten sich die Teilnehmer\*innen insgesamt durch einen hohen Grad an Bildung aus und gehörten fast ausschließlich deutscher Nationalität an (man beachte hier, dass Deutsch als Muttersprache Voraussetzung war). Tab. 2 fasst die Charakteristika der Teilnehmer\*innen der validen Datensätze zusammen.

Tab. 2: Geschlecht, Alter und höchster Abschluss der 514 Teilnehmer\*innen.

<b>Geschlecht</b>	<b>n</b>	<b>Alter</b>	<b>n</b>	<b>Höchster Abschluss</b>	<b>n</b>
weiblich	367	< 20	60	Kein Abschluss	1
männlich	141	20-25	202	Hauptschule	1
anderes	6	26-30	116	Realschule	17
		31-35	37	Abitur	260
		36-40	21	Studium	235
		41-50	33		
		51-60	36		
		> 60	9		

## 2.2 Ergebnisse

Die statistische Auswertung der Ergebnisse der Arousal-, Familiaritäts- und Valenz-Ratings erfolgte mit LibreOffice Calc (The Document Foundation 2020) für grundlegende Kerndaten und mit R (R Core Team 2020) für die weiterführenden Berechnungen.

In einem ersten Schritt wurde die interne Konsistenz der Datensätze berechnet. Das hierzu verwendete Maß des Cronbachschen Alpha gibt Auskunft über die interne Verlässlichkeit der Antworten. Tab. 3 listet die erreichten Werte des Cronbachschen Alpha – da für jede Variable zwei Fragebögen ohne Datenüberschneidungen verwendet wurden, sind die daraus resultierenden Datensätze hier ebenfalls separat gelistet. Zur Vervollständigung wurde dieser Kennwert auch für die Übereinstimmungswerte der obig beschriebenen Präferenz-Erhebung berechnet. Insgesamt weisen alle Datensätze eine exzellente interne Konsistenz auf, was eine fundierte Basis für die weiterführende statistische Analyse bot.

Tab. 3: Cronbachsches Alpha, berechnet für jeden der Fragebögen der Übereinstimmung, Familiarität, Valenz und Arousal.

<b>Fragebogen</b>	<b>n</b>	<b>Cronbachsches Alpha</b>
Agreement 01	81	0,96
Agreement 02	80	0,97
Arousal 01	87	0,98
Arousal 02	86	0,98
Familiarität 01	87	0,99
Familiarität 02	88	0,99
Valenz 01	82	0,96
Valenz 02	84	0,93

Anschließend wurden pro Satzpaar für jede Variable die arithmetischen Mittelwerte, Minimal- und Maximalwerte und die Standardabweichungen berechnet. Die resultierenden Mittelwerte jedes Kontextsatzpaares wurden in die Datenbank MIST übertragen und stehen dort als im Kontext zu verwendende Werte des Arousal, der Valenz und der Familiarität zur Verfügung.

Zusätzlich wurden die arithmetischen Mittelwerte, Minimal- und Maximalwerte und die Standardabweichungen für die Gesamtheit aller Items für die drei Variablen berechnet. Die Ergebnisse, getrennt nach metaphorischen und wörtlichen Ausdrücken, sind in Tab. 4 zusammengefasst.

Tab. 4: Mittelwerte, Standardabweichung, Minimal-, Maximalwerte und Spannweite der Familiarität, Valenz und des Arousal der metaphorischen und wörtlichen Stimuli.

Variable	Metapher					Wörtliche				
	MW	SD	Min	Max	Spannweite	MW	SD	Min	Max	Spannweite
Arousal	3,78	0,94	1,58	6,17	4,60	3,69	0,94	1,64	5,71	4,07
Familiarität	5,95	0,75	2,84	6,93	4,09	6,36	0,58	3,26	6,93	3,67
Valenz	-0,36	1,28	-2,45	2,56	5,01	-0,27	1,28	-2,38	2,41	4,78

Die zugrundeliegenden Verteilungen dieser Werte wurden zusätzlich detailliert untersucht. Zunächst wurde mit dem Shapiro-Wilk-Test auf Normalverteilung geprüft. Der Shapiro-Wilk-Test wurde gewählt, da er mit Datensätzen großen Umfangs in Vergleich zu anderen üblichen Tests am mächtigsten ist (vgl. Razali & Wah 2011). Die Variablen über alle Items hinweg waren nicht normalverteilt, ebenso verhielt es sich für die Verteilungen der Ergebnisse der wörtlichen Ausdrücke. Die Verteilungen der Familiarität und Valenz der metaphorischen Ausdrücke waren ebenfalls klar nicht normalverteilt. Die des Arousal wies im Shapiro-Wilk-Test jedoch uneindeutige Ergebnisse auf, die knapp an der Signifikanzschwelle der Maßwerte saßen. Hier wurde durch die Berechnung der Schiefe der Verteilung jedoch geschlossen, dass die Verteilung leicht rechtssteil beziehungsweise linksschief war und damit nicht streng normalverteilt sein konnte. Die Verteilungen der drei Variablen, aufgeschlüsselt in ihre Gesamtheit und die Unterteilungen in metaphorische und wörtliche Ausdrücke, illustriert Abb. 2.

Sowohl Arousal als auch Familiarität zeigten eine unimodale Form. Die Valenz zeichnete sich durch eine klare bimodale Form aus, deren lokales Maximum im negativen Bereich höher ausfiel als das im positiven. Zudem wiesen Arousal und Familiarität beide eine rechtssteile/linksschiefe Verteilung auf, die im Falle der Familiarität eine deutliche Drängung der Werte im oberen Bereich erfuhr und überaus schmalgipflig ausfiel. Obwohl keine der Verteilungen normalverteilt war, wurden für den Vergleich der Verteilungen der metaphorischen und wörtlichen Stimuli t-Tests durchgeführt. Bei einer Anzahl von jeweils 168 Ausdrücken ist der Stichprobenumfang groß genug, um die Vorannahme der Normalität vernachlässigen zu können (Boneau, 1960; Salkind, 2010). Da alle Berechnungen mit den unveränderten Originalverteilungen durchgeführt werden sollten, wurde zudem auf jegliche Transformation verzichtet.

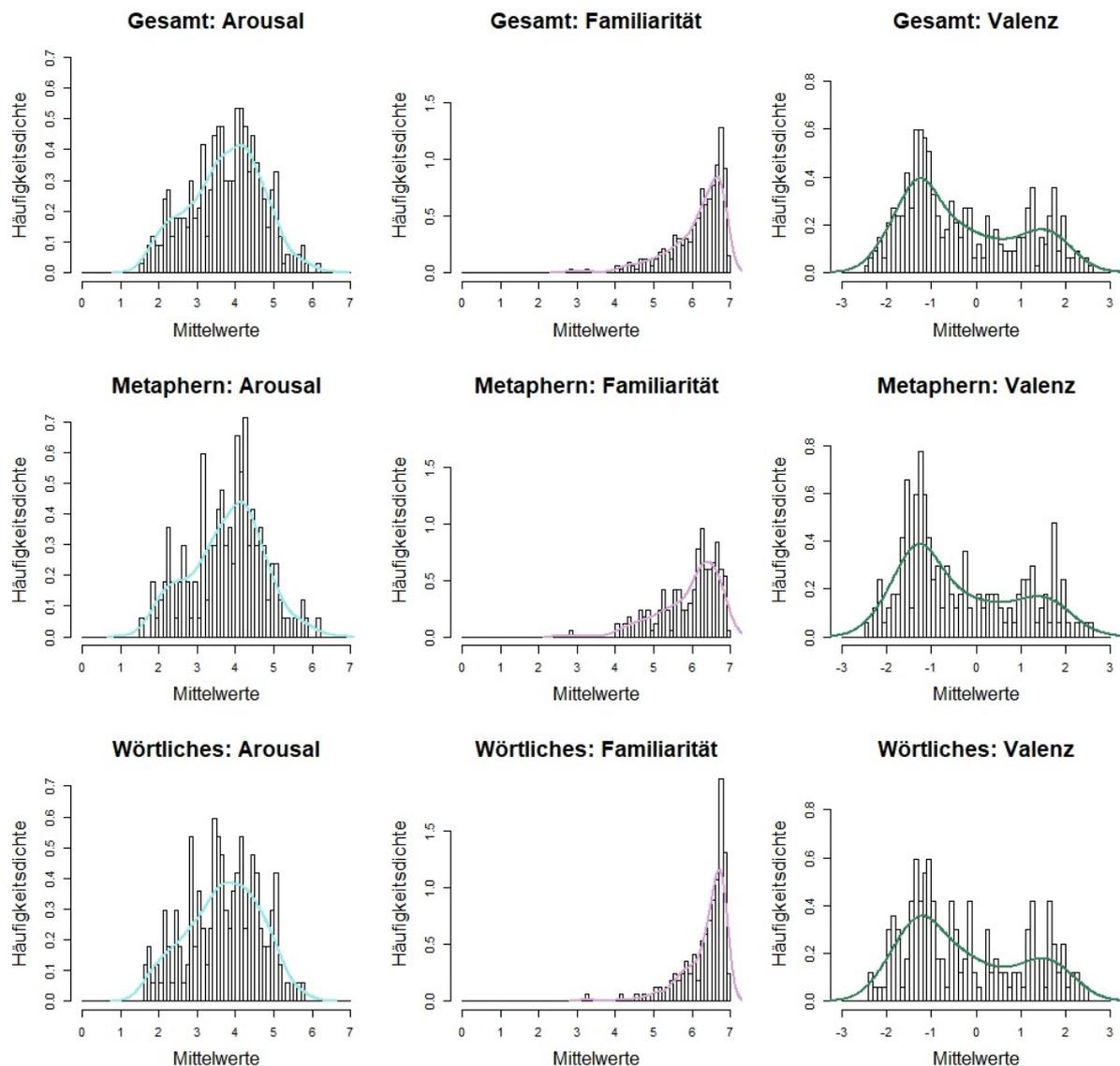


Abb. 2: Verteilungen der Familiarität, Valenz und des Arousal im gesamten Datensatz und den metaphorischen und wörtlichen Stimuli (Histogramme mit Dichtekurve).

Vor der Durchführung der t-Tests ergaben Lavensche Testungen homogene Varianzen in Arousal und Valenz, jedoch eine heterogene Varianz der Familiarität ( $F = 10.266$ ;  $p = 0,0015$ ). Dementsprechend wurden für die ersteren Variablen t-Tests mit homogenen, für die Familiarität t-Tests mit heterogenen Varianzen gerechnet. Die t-Tests wiesen ausschließlich auf signifikante Unterschiede der metaphorischen und wörtlichen Verteilungen in der Familiarität hin ( $t = 5,5768$ ,  $df = 312,73$ ,  $p < 0,001$ ); ein Ergebnis, das auch bei strenger gefassten Konfidenzintervallen signifikant blieb. Die Ergebnisse dieser parametrischen t-Tests wurden durch inhärent nichtparametrische Kolmogorov-Smirnov-Tests validiert.

An die Analyse der isolierten Verteilungen schloss sich in einem nächsten Schritt die Untersuchung der Beziehungen der Variablen zueinander an. Hierfür wurden Korrelationen der drei Variablen, sowohl über alle Ausdrücke als auch separat über metaphorische und wörtliche Stimuli, berechnet. Hinzu kamen

außerdem zwei weitere Variablen, die sich direkt aus den Werten der Valenz ergaben: Valenz<sup>2</sup> und |Valenz|. Valenz<sup>2</sup> stellt laut Citron et al. (2016, 2020) ein Maß für die emotionale Valenz der Stimuli ungeachtet derer positiver oder negativer Ausrichtung dar. Dementsprechend wird die Valenz<sup>2</sup> also als ein Maß der Emotionalität verwendet. An dieser Stelle präsentiert Koller (2021) eine Alternative zu Valenz<sup>2</sup>: |Valenz|. Der Betrag der Valenz hat den Vorteil, dass im Gegensatz zu Valenz<sup>2</sup> keine Verzerrung der Werte relativ zur Null stattfindet. Die Verteilungen der positiven und negativen Werte werden somit realitätsgetreu im "vorzeichenfreien" Bereich abgebildet, wodurch die tatsächlichen Verteilungsmuster und -abstände erhalten bleiben. Zudem erleichtert |Valenz| das Berechnen tatsächlich polynomischer Regressionsmodelle mit sowohl geraden als auch ungeraden Exponenten – Valenz<sup>2</sup> erschwert die Umsetzung polynomischer Modelle mit ungeraden Exponenten.

Korrelationen wurden folglich zwischen den drei Variablen Arousal, Familiarität und Valenz sowie zwischen Arousal und Familiarität mit Valenz<sup>2</sup> und |Valenz| berechnet. Eine Berechnung der Korrelationen zwischen Valenz, Valenz<sup>2</sup> und |Valenz| ist hinfällig. Aufgrund der nichtnormalverteilten Natur der Variablen wurden Spearman-Korrelationen genutzt. Der Spearman-Korrelation liegen Annahmen der Monotonie zugrunde – im Gegensatz zur Pearson-Korrelation, die hier aufgrund ihrer parametrischen Natur nicht verwendet wurde, wird also keine Linearität angenommen. Signifikante Ergebnisse zeigten sich für die Korrelationen zwischen Arousal und Valenz (hoher Zusammenhang), Arousal und Valenz<sup>2</sup> (mittlerer Zusammenhang), Arousal und |Valenz| (mittlerer Zusammenhang), und Familiarität und Valenz (geringer Zusammenhang). Detaillierte Ergebnisse sind in Tab. 5 zusammengefasst.

Tab. 5: Spearman-Korrelationen über alle Items der Variablen Familiarität, Arousal, Valenz, Valenz<sup>2</sup> und |Valenz| (nur Korrelationen mit  $\rho > 0,2$ ).

Korrelation (Spearman)	$\rho$		
	Gesamt	Metapher	Wörtlich
AROUSAL x VALENZ	-0,591 ***	-0,613 ***	-0,563 ***
AROUSAL x VALENZ <sup>2</sup>	0,331 ***	0,316 ***	0,344 ***
AROUSAL x  VALENZ	0,331 ***	0,315 ***	0,344 ***
FAMILIARITÄT x VALENZ	0,250 ***	0,252 ***	0,224 **

\*p < .05, \*\*p < .01 und \*\*\*p < .001

Um eine feinere Einsicht in diese Korrelationen zu gewinnen, wurden partielle Korrelationen untersucht. In diesen partiellen Korrelationen wird auf eine weitere Variable kontrolliert, um auszuschließen, dass diese dritte Variable maßgebend Einfluss auf die Stärke der untersuchten Korrelation hat. Wurde in der Korrelation Arousal x Valenz auf Familiarität kontrolliert, so konnten kaum Veränderungen beobachtet werden (Gesamt:  $\rho = -0,574$  \*\*\*). Gleiches traf auf die Korrelationen Arousal x Valenz<sup>2</sup> (Gesamt:  $\rho = 0,340$  \*\*\*) und Arousal x |Valenz| (Gesamt:  $\rho = 0,340$  \*\*\*) zu. Kontrollierte man jedoch die Korrelation zwischen Familiarität und Valenz auf Arousal, so reduzierte sich Spearmans Rho auf  $\rho = 0,185$  \*\*\* und konnte somit nicht mehr als geringer Zusammenhang eingestuft werden. Die jeweiligen Ergebnisse der Teilmengen der metaphorischen und wörtlichen Untergruppen verhielten sich analog zu denen des

Gesamtdatensatzes. Es verblieben folglich die mittleren bis hohen Korrelationen von Arousal mit den drei Maßen der Valenz.

Da Korrelationen keine Aussage darüber treffen, welchen Einfluss die Variablen aufeinander nehmen, wurden im darauffolgenden Schritt Regressionsanalysen modelliert. Bereits in der Literatur veröffentlichte Studien (Bradley & Lang 1999; Vö et al. 2009; Citron et al. 2016, Müller et al. 2021) berichten von quadratischen Zusammenhängen zwischen Arousal (abhängige Variable) und Valenz (unabhängige Variable). Eine oberflächliche graphische Betrachtung der vorliegenden Verteilungen legte ebenfalls nicht-lineare Zusammenhänge nahe; Linearität konnte auf diese Weise jedoch nicht ausgeschlossen werden. Dementsprechend wurden sowohl lineare als auch polynomische Regressionen für die Zusammenhänge Arousal x Valenz, Arousal x Valenz<sup>2</sup> und Arousal x |Valenz| modelliert, wobei jedes Modell für alle Items und getrennt in die metaphorischen und wörtlichen Kategorien gerechnet wurde. Genauere Analysen dieser Regressionen ergaben detailliertere Untersuchungsschritte, die tiefere Einblicke in die Natur der Zusammenhänge gewährten.

Eine umfassende Schilderung der Regressionsanalysen würde den Umfang des vorliegenden Dokumentes überschreiten. Ihre Ergebnisse sind im Folgenden jedoch kurz zusammengefasst und die vollständige Beschreibung kann in Form der Masterarbeit Sina Kollers (Koller 2021) gerne bei der AG Klinische Linguistik der Universität Marburg angefragt werden.

### 2.3 Diskussion

Für die beschriebene Entwicklung von 168 Stimulisets und die genauere Analyse der bewerteten Variablen bestand durch die durchweg hohen *Agreement*-Werte aller Stimulisets grundsätzlich eine stabile Basis für die darauffolgenden Ratings. *Agreement* korrelierte nicht mit Familiarität – die Vertrautheit mit Ausdrücken erhöht folglich nicht die bewertete inhaltliche Übereinstimmung. Diese Aussage lässt sich nur für die Metaphern und ihre jeweilig als am passendsten bewerteten wörtlichen Entsprechungen im Kontext treffen; für die ausgeschlossenen wörtlichen Entsprechungen liegen keine Familiaritätswerte vor. Die dargelegten Unterschiede der Verteilungen der Variable Familiarität in den metaphorischen und wörtlichen Untergruppen deutet darauf hin, dass wörtliche Ausdrücke signifikant vertrauter als Metaphern sind.

Grundsätzlich wurde die Familiarität als zu bewertender Faktor gewählt, da die Häufigkeit von metaphorischen Bedeutungen nicht gemessen werden kann und somit keine Vorkommenshäufigkeitsdaten vorliegen. Diverse Studien belegten jedoch bereits eine starke Korrelation zwischen Häufigkeit und Familiarität (Rapp 2005; Tanaka-Ishii & Terada 2011). Die Familiarität wurde im empirischen Vorgehen durch zwei Faktoren definiert: erstens die tatsächliche Vertrautheit, d.h. das grundlegende Wissen um die Existenz und die Bekanntheit der Bedeutung des Ausdrucks, und zweitens die subjektive Häufigkeit. Zudem wurden die Familiaritätswerte für die Ausdrücke in Kontext erhoben. Da die Werte für Familiarität, Valenz und Arousal für die kritischen Ausdrücke immer im Kontext der Satzpaare erhoben wurden, wird dringend davon abgeraten, die Ausdrücke und ihre Werte isoliert von diesem Kontext zu verwenden.

Insgesamt rückten durch Spearman-Korrelationen zwei Zusammenhänge in den Fokus: Arousal x Valenz und Arousal x Emotionalität. Die berechneten Regressionsmodelle des Arousals und der Valenz ergaben ein quartisches Modell mit der relativ höchsten Modellgüte, das etwa 46 % der Varianz in Arousal durch Valenz erklärt. Die metaphorischen und wörtlichen Untergruppen unterschieden sich nicht signifikant, was auf eine ähnliche Varianz in den Verteilungen hindeutet. Im Vergleich mit dem Datensatz der sensorischen Stimulisets von Müller et al. (2021) ergibt sich hier grundsätzlich ein deutlich unterschiedliches Bild: Die

Daten des sensorisch-metaphorischen Gesamtdatensatzes präsentierten sich mit deutlich ebenermäßigerer Streuung. Zudem war das passendste Regressionsmodell hier von quadratischer Natur und ergab eine Aufklärung der Varianz in Arousal durch Valenz von bis zu 5,5 %.

Nicht zuletzt durch diese auffälligen Unterschiede wurden die Daten einer detaillierteren Analyse unterzogen. Ein Blick auf die Regressionsgewichte und deren jeweilige Signifikanz des Unterschieds zu Null im tatsächlichen Polynom zeigt, dass nicht extreme Valenzwerte ein deutlich weniger starken Einfluss auf das Arousal ausüben als Extremwerte.

Ein genauere Blick auf die Verteilung der Datenpunkte zeigte, dass im negativen Valenzbereich eine höhere Anzahl der Extremwerte vorliegt als im positiven. Zudem häufen sich im negativen Bereich Datenpunkte, die sowohl eine Valenz  $< -1,5$  als auch Arousal  $> 4$  besitzen. Ausschließlich Ausdrücke, die als negativ bewertet wurden, wurden zusätzlich mit einem Arousal  $> 5,5$  eingestuft. Dies deutet darauf hin, dass negative Ausdrücke, unabhängig von ihrer Bildhaftigkeit, tendenziell höheres Arousal hervorrufen.

Aufgrund der verstärkten Drängung der Werte im extrem negativen Bereich mit hohen Arousalwerten und der Spärlichkeit der extrem positiven Valenzwerte wurde überprüft, inwiefern die Regressionsmodelle von den acht Datenpunkten mit Arousal  $> 4$  und Valenz  $> 2$  beeinflusst wurden. Es zeigte sich, dass ohne diese acht Satzpaare, die lediglich 4,8 % des Gesamtdatensatzes ausmachen, bereits mit einem simpleren quadratischen Regressionsmodell eine fast identische Modellgüte erreicht werden konnte. Spezifischere Diagnostik über Leverage-Diagramme attestierten eine gesunkene Anfälligkeit für Ausreißer dieses Modells. Per Definition sind Extremwerte für die Festlegung hoher Grade der Regressionspolynome verantwortlich. Die Berechnung des Modells ohne die wenigen positiven Extremwerte veranschaulicht dadurch die Anfälligkeit des ursprünglichen quartischen Modells für die Änderung von Extremwerten und die dringende Notwendigkeit von mehr Datenpunkten im extremen Valenz-Bereich. Zudem wurde so illustriert, dass Interpretationen des Zusammenhangs von extrem positiver Valenz mit Arousal nur unter Vorbehalt vorzunehmen sind.

Aufgrund dieser Veränderungen wurden für den positiven und negativen Valenzbereich getrennte Regressionsmodelle für Arousal und Valenz berechnet, die diese Ergebnisse festigten. Hier erlangen quadratische Modelle die höchste Modellgüte: Die negative Valenz erklärt bis zu 38,0 % der Varianz des Arousals, während positive Valenz nur bis zu 22,2 % der Varianz des Arousals erklärt. In einer Zusammenführung der Überlegungen bezüglich der Extremwerte und der Aufteilung des Datensatzes in einen valenzpositiven und -negativen Bereich zeigten Modellberechnungen der Regression Arousal und positiver Valenz ohne die Extremwerte, dass die korrigierten Determinationskoeffizienten in diesen Modellen drastisch sanken – die passendsten Modelle erklären nur noch zwischen 3,2 % und 9 % der Varianz in Arousal durch Valenz. Dies verdeutlicht erneut den unverhältnismäßig starken Einfluss der Extremwerte auf die Modellierung.

Es wurde gefolgert, dass negative Valenz einen stärkeren Einfluss auf die Varianz des Arousals ausübt als positive. Zudem legt die Verteilung des Arousals im positiven Datensatz durch seine leicht linkssteile Form und die geringe Anzahl an Extremwerten nahe, dass positive Ausdrücke tendenziell selten als stark anregend bewertet werden, sondern mehrheitlich im schwächer anregenden Bereich liegen.

Zusammenfassend attestieren die Analysen eine quadratische beziehungsweise quartische Relation zwischen Valenz und Arousal, d.h. Arousalwerte steigen mit sowohl positiv als auch negativ wachsenden Valenzwerten. Diese Ergebnisse stehen im Einklang mit denen der Studien von Bradley & Lang (1999), Citron et al. (2016, 2020), Vö et al. (2009) und Müller et al. (2021). Dies steht jedoch vor einem Kontext von vor allem im positiven und auch nicht-extremen Valenzbereich stark streuenden Daten im vorliegenden

Datensatz und deutlich abweichenden Modellgüten der quadratischen Regressionsmodelle in den Analysen des Datensatzes der sensorischen Stimuli (Müller et al. 2021).

Im Datensatz lässt sich zudem klar erkennen, dass extreme Valenz von extremem Arousal begleitet wird – in den extremen Valenzbereichen streuen die Werte merklich geringer; zudem gibt es keine Ausnahmen dieser Beobachtung. Die Ausprägung dieser Relation ist wiederum davon abhängig, ob die Valenz positiv oder negativ ist. Die grundsätzliche Existenz dieser Relation wird besonders deutlich in den Abbildungen der Emotionalität gegen Arousal, also der Valenz unabhängig der negativen oder positiven Ausprägung.

Diese Relationen legen nahe, dass hohe Emotionalität verlässlich hohes Arousal hervorruft. Eine Gegenüberstellung der graphischen Zusammenhänge zwischen Arousal und Valenz<sup>2</sup> und zwischen Arousal und |Valenz| verdeutlicht außerdem die realitätsgetreue Abbildung der ursprünglichen Valenzwerte im Diagramm der |Valenz|, da hier keine Verzerrung stattfindet und somit weniger Anlass für nicht-realtätsnahe graphische Evaluationen gegeben ist. So bildet das Diagramm der |Valenz| gegen Arousal die breite Streuung der Mehrheit der Werte ( $|Valenz| < 2$ ) deutlicher ab als das Diagramm der Valenz<sup>2</sup> gegen Arousal.

Insgesamt ergaben die Regressionsmodelle der Emotionalität und Arousal eine Varianzaufklärung des Arousals durch Emotionalität von bis zu 11,5 %. Die Regressionsmodelle basierend auf den beiden Maßen der Emotionalität verhielten sich kohärent zueinander: Ein lineares Modell basierend auf Valenz<sup>2</sup> erzielte die beste Modellgüte, während dies für ein quadratisches Regressionsmodell mit |Valenz| der Fall war. Hier war das Regressionsgewicht des Faktors |Valenz| im quadratischen Modell der |Valenz| nicht signifikant von Null verschieden – dennoch verdeutlicht ein solches Polynom die Möglichkeit des Berechnens von polynomischen Regressionsmodellen, die die Emotionalität nicht nur als Komponenten mit geraden Exponenten enthalten. Durch das Einbeziehen ungerader Exponenten werden so feingliedrigere Analysen möglich. Zudem wird so eine inhärente Vorannahme der Valenz<sup>2</sup> vermieden: Durch die Quadrierung der Valenzwerte entsteht eine automatische Vorannahme zur Relation der Valenz und der Emotionalität, die beispielsweise die Berechnung linearer Modelle unnötig aufwändig macht. Dieser mathematischen Voreingenommenheit wird durch das Verwenden der |Valenz| von vorneherein aus dem Weg gegangen. In der Konsequenz wird nahegelegt, als Maß der Emotionalität die |Valenz| anstelle der Valenz<sup>2</sup> heranzuziehen.

## Quellenverzeichnis

- Boneau, C. Alan (1960).** The effects of violations of assumptions underlying the t test. *Psychological Bulletin*, 57 (1): 49–64. <https://doi.org/10.1037/h0041412>
- Bradley, M.M., & Lang, P.J. (1999).** *Affective norms for English words (ANEW): Instructionmanual and affective ratings*. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.
- Citron, F., Lee, M. & Michaelis, N. (2020).** Affective and psycholinguistic norms for German conceptual metaphors (COMETA). *Behavior Research Methods*, 52, 1056–1072. <https://doi.org/10.3758/s13428-019-01300-7>
- Citron, F. M. M., Cacciari, C., Kucharski, M., Beck, L., Conrad, M. & Jacobs, A. M. (2016).** When emotions are expressed figuratively: Psycholinguistic and Affective Norms of 619 Idioms for German (PANIG). *Behavior Research Methods*, 48(1): 91-111. <https://doi.org/10.3758/s13428-015-0581-4>
- The Document Foundation (2020).** LibreOffice 6.4.6.2 [Computer-Software] Berlin Germany.
- Koller, S. (2021).** *Von spannenden Fragen und aufgehenden Lichtern - Eine methodische Meta-Analyse neuronal-bildgebender Studien zur Metaphern- und Idiomverarbeitung mit anschließender Entwicklung von kontrollierten Stimuli für neurolinguistische Studien zur Metaphernverarbeitung im Deutschen* (Masterarbeit, Philipps-Universität Marburg, AG Klinische Linguistik / Neurolinguistik). Unveröffentlicht.
- Leiner, D. J. (2019a).** Too Fast, too Straight, too Weird: Non-Reactive Indicators for Meaningless Data in Internet Surveys. *Survey Research Methods*, 13(3): 229-248. <https://doi.org/https://doi.org/10.18148/srm/2019.v13i3.7403>
- Leiner, D. J. (2019b).** SoSci Survey (Version 3.1.06) [Computer-Software]. <https://www.soscisurvey.de> (Abrufdatum 08.01.2021)
- Leiner, D.J. (2020).** *Zusätzliche Variablen in der Datenausgabe*. SoSci Survey. <https://www.soscisurvey.de/help/doku.php/de:results:variables> (Abrufdatum 08.01.2021)
- Müller, N., Nagels, A. & Kauschke, C. (2021).** Metaphorical expressions originating from human senses: Psycholinguistic and affective norms for German metaphors for internal state terms (MIST database). *Behav Res* (2021). <https://doi.org/10.3758/s13428-021-01639-w>
- R Core Team (2020).** R: A language and environment for statistical computing. R Foundation for Statistical Computing. Wien, Österreich. <https://www.R-project.org/>
- Rapp, R. (2005).** On the relationship between word frequency and word familiarity. In Fisseni, B.; Schmitz, H.-C.; Schröder, B. & Wagner, P. (Ed.), *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen. Beiträge zur GLDV-Tagung 2005 in Bonn* (249–263). Frankfurt: Peter Lang.
- Razali, N. M. & Wah, Y. B. (2011).** Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1): 21-33.
- Salkind, N. J. (2010).** T test, independent samples. *Encyclopedia of research design*, 1, 1552-1556. <https://www.doi.org/10.4135/9781412961288.n475>
- Tanaka-Ishii, K. & Terada, H. (2011).** Word familiarity and frequency. *Studia Linguistica*, 65: 96-116. <https://doi.org/10.1111/j.1467-9582.2010.01176.x>
- Vö, M. L.-H., Conrad, M., Kuchinke L., U. K., Hofmann, M. J. & Jacobs, A. M. (2009).** The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, 41: 534–538. <https://doi.org/10.3758/brm.41.2.534>