



# 100 WORTE

**VON DER WISSENSCHAFT ZUR  
ANWENDUNG - ENTWICKLUNG  
EINER KI-UNTERSTÜTZTEN  
TEXTANALYSE**

<b>Vorwort</b> .....	<b>3</b>
<b>Einführung und Hintergrund</b> .....	<b>3</b>
Wissenschaftliche Grundlage.....	5
Die 100 Worte Kategorien .....	7
<b>Konstruktion</b> .....	<b>10</b>
Wie die 100 Worte Wörterbücher erstellt wurden .....	10
Besonderheiten der 100 Worte Textanalyse .....	13
<b>Psychometrie</b> .....	<b>15</b>
Eigene Arbeiten und Vorgehen.....	15
Psychometrische Eigenschaften der 100 Worte Textanalyse .....	16
<b>Literatur</b> .....	<b>21</b>
<b>Anhang</b> .....	<b>28</b>
Begriffserklärung .....	28



## VORWORT

Die 100 Worte Textanalyse ist das Werk eines Teams bestehend aus Experten unterschiedlicher Fachrichtungen. Jeder dieser Beteiligten liefert wichtige Beiträge zum Gelingen unserer Analyse. In mühevoller Kleinarbeit wurden unzählige Herausforderungen gemeistert, die oftmals ganz neue Lösungen erforderten. All das war nur im Team möglich, da die Komplexität, die unsere Analyse angenommen hat, die Fähigkeiten einer einzelnen Person übersteigen würde. Daher möchte ich diese Gelegenheit nutzen und mich bei allen bedanken, die unsere Analyse möglich gemacht haben und auch weiter daran arbeiten, sie voranzubringen.



## EINFÜHRUNG UND HINTERGRUND

Die Art und Weise, mit der wir bestimmte Worte und Wortkombinationen benutzen, verrät etwas über unsere Persönlichkeit. Diesen Zusammenhang, den viele auch intuitiv vermuten würden, konnte in vielen wissenschaftlichen Studien gezeigt und repliziert werden (Tausczik & Pennebaker, 2010, für einen Überblick). Als Einführung empfiehlt sich auch das Buch „*The Secret Life of Pronouns: What Our Words Say About Us*“ (Pennebaker, 2011).

Die Analyse von Sprache reicht zu den Anfängen der Psychologie zurück. Bereits Sigmund Freud beschäftigte sich 1901 in seinem Werk „Zur Psychopathologie des Alltagslebens“ mit Versprechern und wie diese Aufschluss über Inneres geben können. Rorschach (1921) entwickelte projektive Tests, um über Beschreibungen von mehrdeutigen Tintenflecken auf Motive und Gefühle zu schließen. Beim Aufkommen von Computern war die Analyse von Texten eines der ersten Anwendungsfelder (Mackenzie, Rosenberg, Bergen, & Tucker, 1978). Ebenfalls in dieser Zeit entdeckte J. W. Pennebaker die Bedeutung von Funktionswörtern für die Ermittlung von Persönlichkeit in Texten. In den 1980ern untersucht er zunächst Briefe

von depressiven Patienten und stellte fest, dass Funktionswörter prädiktiv für eine depressive Erkrankung sein können (Pennebaker & Beall, 1986). Diesem ersten Befund folgten weitere und auch andere Wissenschaftler forschten zur Fragestellung, wie Sprache mit Persönlichkeit zusammenhängt. Viele Facetten der Persönlichkeit wurden in Experimenten auf den Zusammenhang zur Sprache untersucht.

Wir können in dieser Arbeit nicht auf jeden Zusammenhang im Detail eingehen. Lediglich der Zusammenhang von Durchsetzungsstärke und Sprache soll hier beispielhaft und zu Anschauungszwecken näher beleuchtet werden.

Wie sich die Durchsetzungsstärke im Speziellen aus der Struktur der Sprache ermitteln lässt, untersuchten u.a. Kacewicz, Pennebaker, Davis, Jeon & Graesser (2014). Die Forscher verglichen die Sprache von Menschen, die entweder einen hohen oder einen niedrigen Status in einer Gruppe inne hatten. Konsistent über verschiedene Experimente hinweg, waren überzufällige Unterschiede in der Verwendung von Funktionsworten (namentlich Personalpronomen) zwischen Vergleichs- und Experimentalbedingung feststellbar.

Menschen mit hohem Status verwendeten weniger die erste Person Singular, aber häufiger die erste Person Plural. Zu diesem Befund kamen einige Jahre zuvor auch schon Sexton und Helmreich (2000). Sie untersuchten die Sprache von Crew-Mitgliedern in einem Simulator.

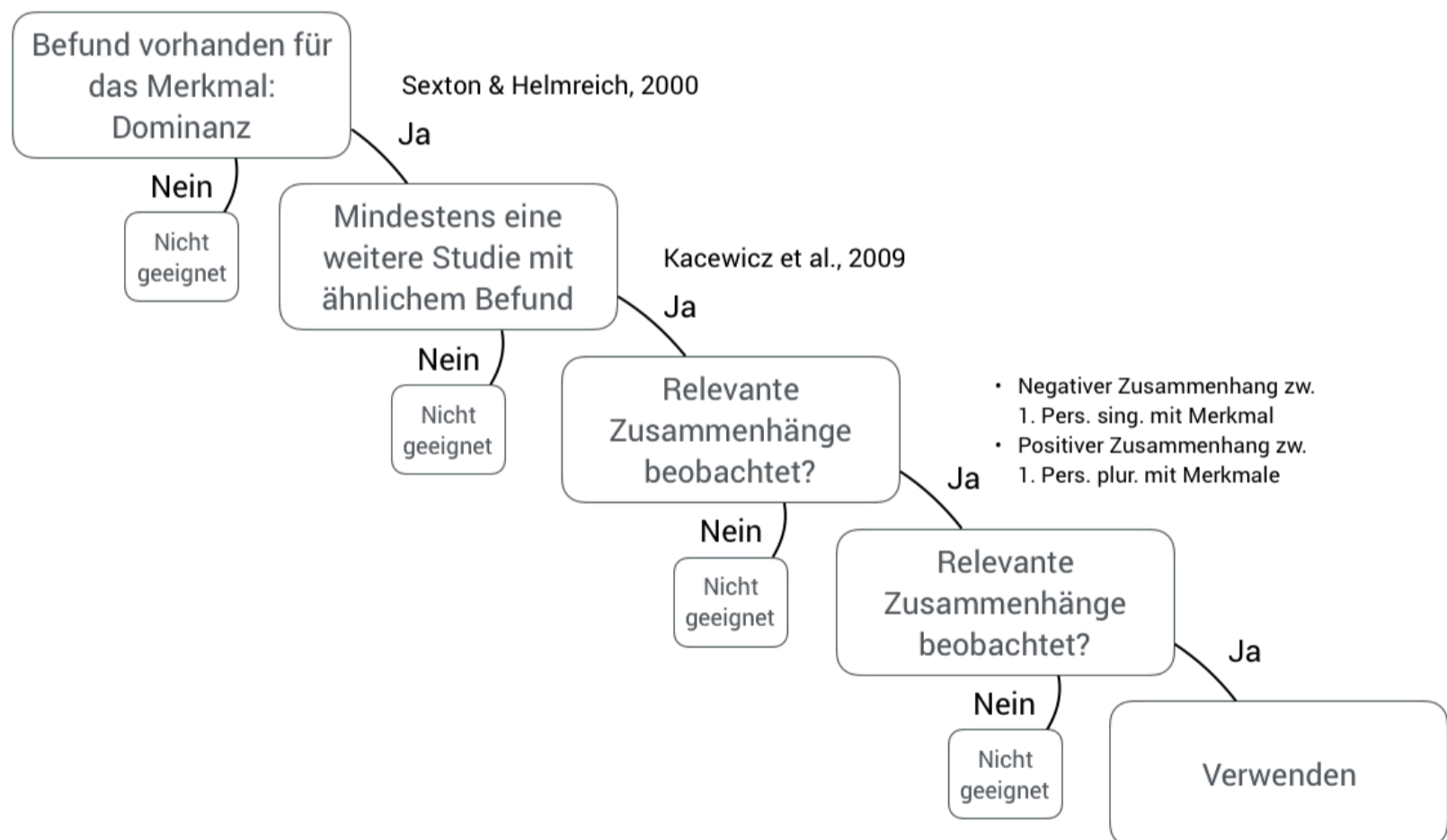
Bei der Entwicklung der 100 Worte Analyse wurden zahlreiche psychologische Studien berücksichtigt, an denen Forscher aus verschiedenen Disziplinen beteiligt waren. In ihren Studien haben die Forscher Hypothesen aufgestellt, wie bestimmte sprachliche Merkmale mit der Persönlichkeit von Menschen in Verbindung stehen könnten. Diese prüften die Forscher in Experimenten und konnten ihre Hypothesen so bestätigen oder widerlegen. Die 100 Worte Textanalyse basiert auf diesen Befunden und wir entwickelten daraus eine regelbasierte, KI-unterstützte Anwendung.

Dieser Beitrag soll einen Überblick darüber geben, (1.) auf welchen wissenschaftlichen Grundlagen die 100 Worte Textanalyse basiert, (2.) wie sie konstruiert wurde und (3.) welche psychometrischen Gütemaße sie aufweist. Die 100 Worte Textanalyse wurde um Methoden des maschinellen Lernens erweitert, wenn Regeln nicht mehr ausreichten, um Text zu verarbeiten. Solche Verfahren sind insbesondere bei der Zuordnung von

Bedeutungen zu einzelnen Worten, deren Morphologie keine eindeutige Bedeutungszuordnung zulässt, notwendig. Zur Verwendung von Lernverfahren erhalten Sie später einen Überblick mit Beispielen, die das Problem verdeutlichen.

## WISSENSCHAFTLICHE GRUNDLAGE

Bevor wissenschaftliche Befunde in die 100 Worte Textanalyse eingeflossen sind, fand jedoch eine kritische Prüfung statt. Das war nötig, denn nur so konnte eine hohe Qualität sichergestellt werden. Befunde wurden für unsere Anwendung ausgewählt, wenn sie einen relevanten Beitrag zur Erklärung eines Persönlichkeitsmerkmals leisteten (= statistisch signifikanter<sup>2</sup> Zusammenhang zwischen Sprachelement und Persönlichkeitsmerkmal mit einem Korrelationskoeffizienten min. größer gleich .20) und sich in mindestens zwei Studien nachweisen ließen. Für das bereits oben eingeführte Merkmal Durchsetzungsstärke / Dominanz ist das Vorgehen beispielhaft dargestellt.



Wurden relevante Sprachindikatoren für Persönlichkeitsmerkmale gefunden, überträgt ein Team bestehend aus Psychologen, Data-Scientists und Programmierern diesen Zusammenhang in Code-Sprache, sodass sie im 100 Worte Core nutzbar werden. Eine

Auflistung aller aus Sprache gemessener Merkmale findet sich im Anhang (siehe Überblick Studien). Die folgende Tabelle enthält eine Beschreibung unserer Sprachindikatoren und wissenschaftliche Befunde, die die Verwendung der Kategorien rechtfertigen.

Tabelle 1. 100 Worte Kategorien und Befunde

100 Worte Kategorie	Beispiele	# Wörter	Psychologisches Korrelat	Publizierte Artikel	Eigene Artikel
<b>Personalpronomen</b>	Ich, ihr, sein	57	Sozialer Bezug	1,2	
<b>Erste Person singular</b>	Ich, mein, mir	9	Psychisches Wohlbefinden, Ehrlichkeit, sozialer Status, informell	3,4,5,6,7,8,9,10,11,12,13,14, 15,16,17,18,24,29,32	-
<b>Erste Person plural</b>	Wir, uns, unser	8	Sozialer Status, soziale Verbundenheit	19,20,21,28,29,32	-
<b>Dritte Person singular</b>	Er, sie, ihrem	14	Soziales Interesse, Unterstützung	22,23,29,32	-
<b>Artikel</b>	Das, dem, ein	18	Analytisch, Interesse an Dingen	24,25,30	-
<b>Vergangenheit</b>	aufgehört, durfte, gewesen	426	Vergangenheitsperspektive	26,27,32	-
<b>Zukunft</b>	ahne, erzielt, plane	222	Zukunftsperspektive, Zielorientierung	33,57	-
<b>Präposition</b>	zum, unter, für	267	Bedacht auf Präzision	1,30	-
<b>Adverbien</b>	Alle, dann, genau	1027	Authentizität	2	-
<b>Hilfsverben</b>	Bin, habe, ist	117	-	-	-
<b>Negationen</b>	kein, nicht, niemand	22	Inhibition	24,31	-
<b>Schimpfworte</b>	Aas, Scheiß, Trottel	2673	Aggression	12,32	-
<b>Positive Emotion</b>	Adrett, Freude, begabt	7739	Positive Stimmung	32,33	-
<b>Negative Emotion</b>	Fehler, freudlos, misstrauisch	15472	Negative Stimmung	24,32	-
<b>Angst</b>	ängstlich, erstarrt, sorgenvoll	1026	Ausdruck von Angst	34,35	-
<b>Ärger</b>	Ärgern, blöd, erzürnt	2487	Ausdruck von Ärger	32,34,35,36	-
<b>Trauer</b>	Apathie, düster, hilflos	1365	Ausdruck von Trauer	34,35,36	-
<b>Diskrepanz</b>	könnte, möglicherweise, vielleicht	941	Authentizität, Ehrlichkeit	24,58	-
<b>Absolutheit</b>	Absolut, muss, zweifellos	1446	Authentizität, Promotion Focus	58,59	-
<b>Machtmotiv</b>	Applaus, unterdrücken, führen	4314	Motiv nach Macht, Führung, Status, Kontrolle	40,41,42,43,44,45,46,48	54,55

100 Worte Kategorie	Beispiele	# Wörter	Psychologisches Korrelat	Publizierte Artikel	Eigene Artikel
<b>Leistungsmotiv</b>	Ambition, glücklos, Ziel	3074	Motiv nach Leistung, Weiterentwicklung, Vorankommen	40,41,42,43,44,45,46,47,48	54,55
<b>Beziehungsmotiv</b>	Beisammen, fremd, uns	2044	Motiv nach guten Beziehungen	40,41,42,43,44,45,46,48	54,55
<b>Gewinn</b>	Abenteuer, Bonus, Reiz	1739	Fokus auf Gewinn, Promotion Focus	49,50,51	-
<b>Verlust</b>	Aversion, Furcht, Kosten	1287	Fokus auf Verlust, Prevention Focus	49,50,51	-
<b>Maskulin</b>	Analytisch, direkt, Gewalt	1359	Ausdruck von Männlichkeit, Dominanz, Stärke	37,38,39,53	56
<b>Feminin</b>	Anvertrauen, engagiert, helfen	1188	Ausdruck von Weiblichkeit, Verbundenheit, Milde	37,38,39,53	56
<b>Descriptive Action Verbs</b>	backen, blühen, gießen	3668	Ausdruck von Konkretheit	52,60	-
<b>State Verbs</b>	Anlegen, besitzen, ist	551	Ausdruck von Abstraktion	52,60	-
<b>Sozialer Status</b>	-	-	-	21,30	-
<b>Denkstil</b>	-	-	-	31	-
<b>Emotionalität</b>	-	-	-	62	-
<b>Regulatorischer Fokus</b>	-	-	-	49,50,51	-

## DIE 100 WORTE KATEGORIEN

Die 100 Worte Sprach-Kategorien umfassen verschiedene psychologische Merkmale. Insgesamt beinhaltet unsere Textanalyse 38 Kategorien, die sich in drei Gruppen aufteilen: Merkmale der Persönlichkeit, Merkmale der Stimmungen und Emotionen und Merkmale der Motive und Bedürfnisse. Diese Merkmale lassen sich mithilfe einer Mensch-Metapher eingängig darstellen. Eine Darstellung und Beschreibung aller Kategorien findet sich in Tabelle 2.

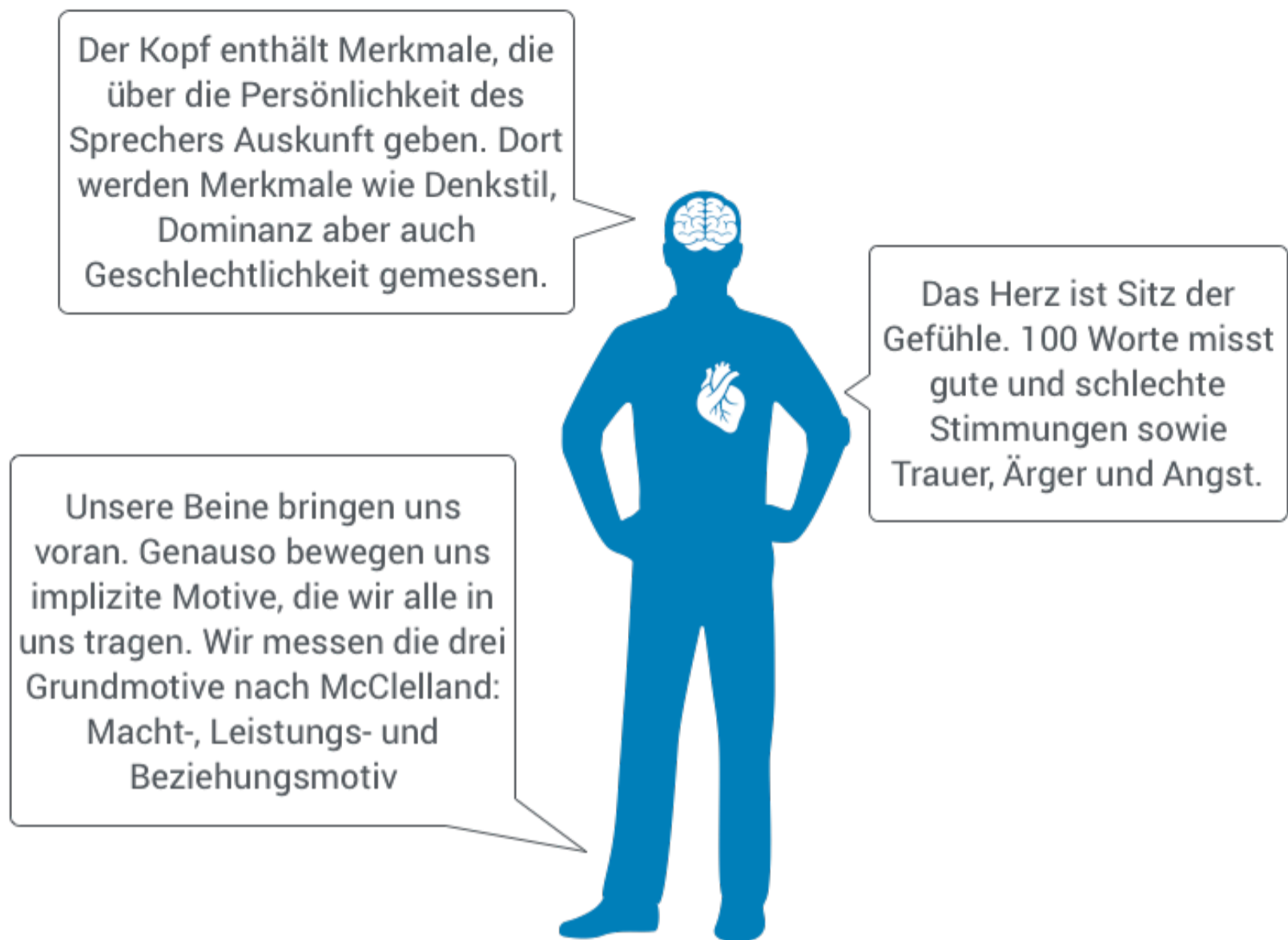


Tabelle 2. 100 Psychologische Konzepte und zugrundeliegende Sprachmerkmale

Psychologisches Konzept	Kategorie bedeutet...	Kategorie beinhaltet...
<b>Stimmungen und Emotionen</b>	Stimmungen und Emotionen geben Auskunft über die generelle Stimmung sowie die aktuelle Befindlichkeit. Indikative Worte für gute bzw. schlechte Stimmung sind z. B. „freudig“, „sorgenfrei“, „gefällt“; „unzufrieden“, „gelangweilt“, „ignorant“.	<ul style="list-style-type: none"> <li>• Positive Emotion</li> <li>• Negative Emotion</li> <li>• Angst</li> <li>• Ärger</li> <li>• Trauer</li> </ul>
<b>Bedürfnisse und Motive</b>	Menschen haben unterschiedliche Motive und Bedürfnisse, die ihr Handeln und Denken (damit auch das Sprechen) leiten. Besonders sympathisch und bezogen wirkt das Beziehungsmotiv in der Sprache, weil hier Worte der Verbundenheit (z. B. "gemeinsam", "zusammen", "miteinander") verwendet werden.	<ul style="list-style-type: none"> <li>• Machtmotiv</li> <li>• Leistungsmotiv</li> <li>• Beziehungsmotiv</li> </ul>
<b>Charakteristika</b>	Hierunter sind verschiedene Kategorien zusammengefasst, die die Persönlichkeit beschreiben.	<ul style="list-style-type: none"> <li>• Denkstil: Analytisch vs. intuitiv</li> <li>• Dominanz: dominant vs. nachgiebig</li> <li>• Authentizität: formell vs. frei-heraus</li> <li>• Maskulin / Feminin</li> </ul>



## Psychologisches Konzept

## Kategorie bedeutet...

## Kategorie beinhaltet...

<b>Regulatorischer Fokus</b>	Der regulatorische Fokus beschreibt zwei grundlegende motivationale Orientierungen. Menschen mit einem Verlustfokus ist es wichtig, Fehler und Gefahren zu vermeiden. Bei einem Gewinnfokus steht dagegen das Wachstum und der Gewinn im Vordergrund.	<ul style="list-style-type: none"><li>• Gewinn</li><li>• Verlust</li></ul>
<b>Zeitliche Perspektive</b>	Menschen haben unterschiedliche zeitliche Orientierungen in der Sprache. Manche verwenden eher eine zukunftsorientierte Sprache, andere eine vergangenheitsorientierte.	<ul style="list-style-type: none"><li>• Vergangenheit</li><li>• Zukunft</li></ul>
<b>Sprachliche Position (aktiv vs. passiv)</b>	Menschen, die aktiv formulieren, leiten das Gespräch. Menschen, die passiv formulieren, sind Empfänger von Anweisungen. Sie werden geleitet. Im Umgang mit Kunden empfiehlt sich eine aktive Sprache.	Agent: Aktive Position Patient: Passive Position
<b>Sprachliche Genauigkeit</b>	Sprache unterscheidet sich in konkrete und unkonkrete Sprache. Ist eine Sprache unkonkret, lässt sie Raum für Unklarheit und Interpretation. Im Kundenkontakt ist eine konkrete Sprache anzustreben.	DAV: Descriptive Action Verbs SV: State Verbs
<b>Sicherheit / Unsicherheit</b>	Wenn sich Menschen einer Sache sicher sind, drücken sie das auch in absoluten Worten (z. B. "Immer", "auf jeden Fall") aus. Andernfalls verwenden sie Worte der Unsicherheit (z. B. "Möglicherweise").	<ul style="list-style-type: none"><li>• Absolutheit</li><li>• Diskrepanz</li></ul>
<b>Persönliche Sprache</b>	Sprache kann persönlich oder unpersönlich wirken. Bei einer persönlichen Ansprache werden Menschen direkt angesprochen, indem sie miteinander, übereinander oder über sich selbst sprechen.	<ul style="list-style-type: none"><li>• Personalpronomen</li><li>• Erste Person singular</li><li>• Erste Person plural</li><li>• Dritte Person singular</li></ul>
<b>Funktionsworte</b>	Diese Kategorien haben keine inhaltliche Bedeutung, geben der Sprache aber Struktur. Sie sind wichtig, um Denkstile, Dominanz oder Authentizität in der Sprache zu messen.	<ul style="list-style-type: none"><li>• Personalpronomen</li><li>• Artikel</li><li>• Präpositionen</li><li>• Pronomen</li><li>• Adverbien</li><li>• Hilfsverben</li></ul>
<b>Inhaltskategorien</b>	Diese Kategorie umfasst Worte, die einen bestimmten Inhalt ausdrücken.	<ul style="list-style-type: none"><li>• Finanzen</li></ul>



# KONSTRUKTION

Von Daniel Spitzer und Simon Tschürtz

Die Grundlage einer jeden regelbasierten Textanalyse sind sog. Corpora, die 1- oder Mehr-Gramme enthalten und die meistens einem Thema (z. B. Linguistische / psychologische Kategorie) angehören. Im folgenden Kapitel soll ein Überblick über die Konstruktionsweise der 100 Worte Wörterbücher gegeben werden.

## WIE DIE 100 WORTE WÖRTERBÜCHER ERSTELLT WURDEN

### 1. Welche Merkmale sind interessant?

Wie oben beschrieben, fand zunächst eine ausführliche Recherche zu relevanten Forschungsbefunden zu Zusammenhängen zwischen psychologischen Merkmalen und sprachlichen Markern statt. Dies war nötig um entscheiden zu können, ob und welche sprachlichen Marker mit psychologischen Merkmalen zusammenhängen. Tabelle 1 gibt Auskunft, welche Studien herangezogen wurden, um die Verwendung der dort aufgeführten Sprachmerkmale zu rechtfertigen.

### 2. Erstellung der Wortlisten

Funktionswort-Kategorien: Funktionsworte sind für die Konstruktion (Grammatik) eines Satzes wichtig. Sie haben selbst keine inhaltliche Bedeutung, sondern bilden ein Gerüst für Inhaltsworte. Wie in Tabelle 2 zu sehen ist, weisen Funktionsworte einige Zusammenhänge zu psychologischen Merkmalen (wie z. B. zum Sozialen Status oder zur psychischen Gesundheit) auf. Da es allgemein gültige Definitionen gibt, was Funktionsworte sind und in welche Kategorien sie zu unterteilen sind, war es nicht schwer, eine Sammlung eben dieser Funktionsworte-Kategorien anzulegen. Zur Bildung der Kategorien und zur Einteilung von Wörtern verwendeten wir das Stuttgart-Tübingen Tagest (Schiller, Teufel, Stöckert & Thielen, 1999), das Wörter zu 54 Tags zuordnet.

Inhaltswort-Kategorien: Da für die Inhaltswort-Kategorien keine allgemein gültige Definition besteht, war das Vorgehen zur Erstellung dieser Kategorien komplexer als für die Funktionswort-Kategorien. Das Vorgehen bestand hauptsächlich aus drei Phasen, die sich in manuelle und Data Science Methoden gliedern lassen. Am Anfang stand eine Initialisierungsphase. Hier machten sich verschiedene Personen Gedanken darüber, welche Worte mit einem bestimmten psychologischen Konstrukt assoziiert sind. Diese Assoziationen wurden gesammelt und dienten als Grundlage für alle weiteren Schritte. In der zweiten Phase wurden durch ein Wort-Ähnlichkeitsverfahren (=LSA) ähnliche Worte zu den bereits bestehenden gesammelt. Diese Verfahren stellen Wörter in Vektoren dar, welche in einem Vektorraum aufgespannt werden. Durch eine darauf aufbauende Clusteranalyse werden Wörter gefunden, die in einem ähnlichen Sinn vorkommen. Im dritten Schritt wurde die Basisraten-Häufigkeit aller gefundenen Wörter erhoben und diese nach Häufigkeit sortiert. Die Wörter mit der größten Häufigkeit standen ganz oben, die mit der niedrigsten, ganz unten. Im dritten Schritt wurde die interne Konsistenz der erstellten Wörterbücher berechnet und durch schrittweise Wegnahme einzelner Wörter untersucht, welchen Beitrag sie zur gesamten Konsistenz leisten.

### **3. Bewertung der erstellten Wörterbücher**

Nachdem die vorläufigen Wörterbücher erstellt waren, musste entschieden werden, welche Wörter endgültig aufgenommen werden. Bei dieser Entscheidung halfen uns die oben beschriebenen statistischen Kenngrößen wie die Ähnlichkeit, die Basisrate und der Beitrag zur Konsistenz. Es wurden Schwellenwerte für die genannten Kenngrößen formuliert und solche Wörter von den Listen entfernt, die nicht in die Listen passten. Am Ende wurden die Listen drei Personen vorgelegt, die entscheiden sollten, ob die darin enthaltenen Wörter in die Kategorie passen. Es blieben die Worte übrig, bei denen mindestens zwei Annotatoren zu einer Zustimmung kamen. Die Wörter, für die keine eindeutige Zuordnung durch die Annotatoren möglich war, wurden danach separat untersucht, da festgestellt wurde, dass es sich dabei häufig um ambige Wörter gehandelt hat.

Ambige Worte: Ambige Worte sind solche, die mehrere Bedeutungen tragen können. Ein Beispiel für ein solches ambiges Wort ist „nötigen“. Es trägt im Kontext „Ich nötige dich“ eine Machtkomponente. Im Kontext „schicken Sie bitte die nötigen Unterlagen“ hat das Wort diese Bedeutung allerdings nicht. Wenn „nötigen“ in das Wörterbuch „Macht und Führung“ aufgenommen werden würde, dann würde die Textanalyse das Wort in allen Kontexten als Machtwort klassifizieren, obwohl es nicht immer eines ist. Schwartz, Eichstaedt, Blanco, Dziurzyński, Kern et al., (2013) untersuchten die häufigsten Fehler, die durch regelbasierte Textanalysen gemacht wurden, darunter die Fehlklassifizierung

komplexer Negationen, Ironie, falsche Autorenschaft, Rechtschreibfehler und am häufigsten, die Verwendung falscher Bedeutungen. Bei der Erstellung der Wörterbücher hatten wir also die Option, ambige Worte nicht aufzunehmen (um dadurch keine false-positives zu produzieren), sie aufzunehmen (um keine false-negatives zu produzieren) oder einen Weg zu finden, den Kontext eines Wortes zu berücksichtigen, um die Bedeutung des Wortes richtig zu antizipieren. Da wir den Anspruch haben, ein bestmögliches Ergebnis zu erzielen, haben wir uns schließlich dafür entschieden, die Wörter dann aufzunehmen, wenn wir sicher sein können, dass weder Fehler erster, noch zweiter Art begangen werden. Dazu nutzen wir ein maschinelles Verfahren, das den Kontext eines Wortes berücksichtigt. Um dieses Verfahren, welches wir unter dem Begriff „Word-Sense Counting“ zusammenfassen, haben wir unsere Analyse ergänzt.

Im Detail gehen wir folgendermaßen vor: Zur Disambiguierung von Wörtern, die sich anhand ihres Part-of-Speech (POS: Die Zuordnung von Wörtern eines Textes zu ihren Wortarten) unterscheiden lassen, verwenden wir das sog. POS-Tagging Verfahren. Diese Verfahren sind Teil des Natural Language Processing. Beispiel für diese Art von ambigen Wörtern ist „nötigen“. Im oben schon vorgestellten Kontext „Ich benötige dich“, wird es als Verb, im Kontext der Aufforderung „schicken Sie bitte die benötigten Unterlagen“, als Adjektiv gebraucht. Es gibt Software-Pakete, die entsprechende Analysen anbieten. Dazu zählen z. B. Stanford CoreNLP (<https://stanfordnlp.github.io/CoreNLP/>) und spaCy (<https://spacy.io>). Wir verwenden spaCy, weil es gegenüber Stanford CoreNLP einen entscheidenden Geschwindigkeitsvorteil aufweist, wie wir durch eigene Benchmark-Tests herausfanden. Es gibt allerdings auch ambige Worte, die sich nicht durch ihren POS unterscheiden lassen - „Leiter“ ist ein Beispiel hierfür. In den Sätzen „Stell mir bitte die Leiter an die Wand“ und „Die Leiter sind gerade zu Tisch“, wird „Leiter“ in beiden Fällen als Nomen verwendet. Sie sind folglich nicht mehr durch ihren POS zu unterscheiden. Für Menschen ist die Bedeutung der beiden „Leiter“ einfach zu erfassen. Für Maschinen stellt es allerdings eine schwierige Aufgabe dar. Um auch solche Worte in unsere Wörterbücher aufnehmen zu können, bedarf es eines Verfahrens, das den Kontext der in Frage stehenden Wörter berücksichtigt, um eine Einschätzung über die Wortbedeutung zu erhalten. Eine Möglichkeit den Kontext von bestimmten Wörtern zu berücksichtigen, besteht in den sogenannte Embedded Language Models. Diese gehen noch einen Schritt weiter, als die unter Verfahrensschritt 2 genannten Word Embeddings, indem Sie alle Wörter eines Textes mitberücksichtigen. In den genannten Beispielen mit den unterschiedlichen Bedeutungen von „Leiter“, errechnet dieses Verfahren unterschiedliche Richtungen von Vektoren. Da wir die für uns relevanten ambigen Worte bereits festgelegt hatten, konnten wir nun Modelle trainieren, die die richtige Bedeutung des jeweiligen Wortes identifizierten. Dabei wurden diese Worte im Kontext (=Beispielsätze) drei Annotatoren vorgelegt, die entscheiden mussten, ob beispielsweise

das Wort „Leiter“ in dem Satz mit dem intendierten Sinn vorkommt. Nach ca. 15 Sätzen pro Wort, hat das System gelernt, den richtigen Sinn zu verstehen.

Indem wir in der Lage waren, Wortbedeutungen zu berücksichtigen, konnten wir nun zusätzlich ambige Worte in unsere Wörterbücher aufnehmen. Die Kombination eines regelbasierten Ansatzes gepaart mit den hier beschriebenen Methoden des maschinellen Textverständnisses machen unsere Analyse einzigartig und sorgen für eine höchstmögliche Validität und Objektivität.

## **BESONDERHEITEN DER 100 WORTE TEXTANALYSE**

Neben den oben beschriebenen Merkmalen unseres Verfahrens, haben wir weitere Fehlerquellen und Besonderheiten von Textverfahren identifiziert und jeweils Lösungen mit einer Mischung aus maschinellen Lernen und regelbasierten Verfahren gefunden.

Partikelverben: In der deutschen Sprache werden Verben, die aus zwei Teilen bestehen und deren Anordnung je nach Satzkonstellation unterschiedlich ausfällt, häufig verwendet. Beispiel hierfür ist das Partikelverb „zusammenkommen“. Es kann sowohl in der dargestellten Form vorkommen als auch in der modifizierten Form, wenn das Präfix nach dem Suffix im Satz erscheint („Vielleicht kommen wir beide bald wieder zusammen“). Die Erfassung dieser sog. Partikelverben stellt eine besondere Anforderung an ein Textanalyseprogramm dar. Die Schwierigkeit besteht darin, die über einen Satz verteilten Elemente des Partikelverbs zu identifizieren. Hierbei half uns das schon oben erwähnte NLP-Programm spaCy. Auch hier verwendeten wir die Analyse des POS. Mithilfe des POS-Tagging kann auch ermittelt werden, ob zwei Elemente in einem Satz zu einem Partikelverb gehören. Indem wir auch Partikelverben identifizieren können, müssen wir in unserer Analysen nicht auf diese verzichten und fügten Partikelverben zu unseren Wörterbüchern hinzu.

Wortstamm-Bildung: Bei vielen regelbasierten Textanalysen ist es üblich (z. B. LIWC), die verschiedenen Formen eines Wortes (z. B. die Deklinationen des Verbs „gehen“) nicht ins Wörterbuch aufzunehmen, stattdessen alle Formen auf einen gemeinsamen Wortstamm zu kürzen um so nur diesen Stamm zu speichern. Stemming hat den Vorteil, dass Wörterbücher deutlich kürzer werden und die Verarbeitung, gerade bei großen Textmengen, effizienter wird. Aus der Arbeit von Schwartz et al., (2013) wissen wir aber, dass dadurch die Rate von Missklassifikationen ansteigt. Dieser Fall soll mit dem Beispiel „Note“ verdeutlicht werden. Durch Stemming mit dem NLTK-Stemmer (<http://www.nltk.org/>

[howto/stem.html](#)) wird „Note“ auf „Not“ gekürzt und so zu einem ambigen Wort, da „Not“ auch eine andere Bedeutung trägt und in unserer negative Emotions-Kategorie enthalten ist. Wir erachten den beschriebenen Nachteil, der Stemming mit sich bringt, als gewichtiger als die Vorteile und verwenden für unsere Analyse daher kein Stemming.

Verstärker-Präposition: Viele Adjektive haben eine verstärkende Präposition wie z. B. „sehr schön“. Diese Verstärker betonen eine Sache besonders und sollten daher auch in einer Textanalyse berücksichtigt werden. Die 100 Worte Textanalyse erfasst diese Verstärker und gibt der zum darauffolgenden Wort gehörigen Kategorie ein stärkeres Gewicht. Ohne Verstärker zählen alle gefundenen Worte und Formulierungen einfach. Mit Verstärker werden sie 1,2 fach gezählt.

Negationen: Einige Kategorien in der 100 Worte Textanalyse drücken eine Wertung aus. Mit diesen Kategorien ist es möglich, Polaritäten aufzuzeigen und Sachverhalte positiv oder negativ darzustellen. Gerade die Emotionskategorien eignen sich hierfür sehr gut. Jedoch kann jede Polarität durch Verneinung ins Gegenteil verkehrt werden. So wird das Wort „schön“ durch die Verwendung einer Negation „nicht schön“ bedeutungsmäßig ins Gegenteil verkehrt. Um solche Umkehrungen zu erfassen, muss unsere Textanalyse Verneinungen erkennen. Dazu haben wir ein Verfahren entwickelt, dass die häufigsten Formen von Verneinungen (= Negation erscheint direkt „nicht schön“ oder unmittelbar „nicht sehr schön“ vor der in Frage stehenden Valenz) erfasst werden. Komplexe Verneinungen wie etwa „Nie sind die Herrschaften zufrieden“, bei der die Negation „Nie“ weit entfernt von der Valenz „zufrieden“ erscheint, werden dagegen nicht erfasst. Die Auftretenswahrscheinlichkeit solcher komplexer Negationen ist - unserer Einschätzung zufolge - zu gering, um den Aufwand zur Implementierung einer solchen Erfassung zu rechtfertigen.

Ngramme: Ngramme sind Wortkombinationen, die aus zwei oder mehr Worten bestehen. Das „N“ in „Ngramme“ dient dabei als Platzhalter für die tatsächliche Anzahl der in der Kombination vorkommenden Worte. Die Erfassung von Ngrammen ist wichtig, weil sich Bedeutungen manchmal nur in der Kombination von Worten ausdrücken. Im Satz „Mit dir gehe ich durch dick und dünn“ drückt das 4-Gramm „durch dick und dünn“ eine Bedeutung (z. B. Beziehung) aus, die die einzelnen Substituenten alleine oder in einer anderen Kombination nicht hätten. So würde es keinen Sinn machen, „durch“, „dick“, „und“ und „dünn“ als einzelne Worte in unserem Beziehungswörterbuch zu speichern, weil nur die Kombination die spezifische (Beziehungs-)Bedeutung trägt. Wir sammeln folglich Ngramme und fügen sie unseren Wörterbüchern hinzu. Dabei unterscheidet sich das Vorgehen zur Identifikation von Ngrammen zum oben beschriebenen Ansatz zur Erstellung von

Wörterbüchern. Ngramm-Kandidaten werden gebildet, indem die häufigsten Kombinationen von 2-/3-/...-Grammen aus einem gegebenen Datenset mithilfe automatischer Verfahren extrahiert und nach Auftretenshäufigkeit sortiert werden. Diese Listen werden so gekürzt, sodass nur Ngramme übrig bleiben, deren Auftretenshäufigkeit über einem Prozent liegt. Anschließend bewerten drei Annotatoren diese und es bleiben solche Ngramme übrig, die eine Mehrheit als dazugehörig klassifiziert hat.

Zusammengesetzte Kategorien: Wie bereits erwähnt, werden einige der 100 Worte Kategorien zusammengefasst, um über sprachpsychologische Merkmale Aufschluss zu geben. Diese Merkmale sind: Sozialer Status, Denkstil, Emotionalität, Regulatorischer Fokus und Authentizität. Die genaue Kombination der Kategorien zu den genannten Merkmalen ist der in Tabelle 1 aufgeführten Literatur zu entnehmen.



# PSYCHOMETRIE

## EIGENE ARBEITEN UND VORGEHEN

100 Worte gibt sich aber mit der bloßen Übertragung von bestehenden Forschungsbefunden nicht zufrieden. Unser Anspruch ist es, die Qualität unserer Textanalyse selbst nachzuweisen. Um das zu erreichen, führen wir Studien durch und vergleichen unsere Analyse mit anderen Verfahren sowie mit Wettbewerbern. Aus dieser Forschungstätigkeit sind bereits vielversprechende Resultate entstanden. Wir verglichen dabei die 100 Worte Stimmungskategorien mit drei gängigen Sentiment-Corpi für die deutsche Sprache: SentiWS, German Polarity Clues und Zürich Lexicon. Wir ließen die vier Wörterbücher in der Vorhersage von Stimmungen in rund 8000 deutschen Tweets gegeneinander antreten. Die 100 Worte Stimmungskategorien erzielten dabei eine vier bis 15 Prozent höhere Genauigkeit (=F-Score<sup>1</sup>) in der Vorhersage der Stimmung gegenüber den konkurrierenden Wörterbüchern.

Wir evaluierten aber nicht nur unsere Stimmungs- sondern auch unsere Motivkategorien. Dazu verwendeten wir einen öffentlich zugänglichen Datensatz von Schultheiss, der 2009,

im Rahmen der Studie „*Activity inhibition: a predictor of lateralized brain function during stress*“, entstand: <http://www.psych2.phil.uni-erlangen.de/%7Eoschult/humanlab/publications.htm>

Die Details der Studie sollen hier nicht Thema sein. Interessierte können sie hier einsehen: [https://www.100worte.de/wp-content/uploads/2019/07/100Worte\\_Validierung\\_Motive.pdf](https://www.100worte.de/wp-content/uploads/2019/07/100Worte_Validierung_Motive.pdf)

Diesen Befund konnten wir mit einem weiteren Daten von S. Dörr (A47 Consulting) replizieren und die Ergebnisse bestätigen. Eine Veröffentlichung dazu ist gerade in Arbeit und soll im Herbst 2019 erscheinen.

## **PSYCHOMETRISCHE EIGENSCHAFTEN DER 100 WORTE TEXTANALYSE**

In diesem Abschnitt geht es um die Reliabilität. Hierunter ist die Zuverlässigkeit, also die Stabilität der Messwerte über die Zeit, zu verstehen. An die Beurteilung von Zuverlässigkeit von Textanalyseverfahren sind besondere Herausforderungen gestellt, wie bereits Pennebaker, Boyd, Jordan, & Blackburn (2015) und Meier, Boyd, Pennebaker, Mehl, Martin et al. (2018) festhalten. Im Folgenden soll auf die besonderen Herausforderungen eingegangen werden.

### **Textanalyse und Reliabilität**

Für fragebogen-basierte Tests ist das Vorgehen zur Evaluation der Zuverlässigkeit klar und vergleichsweise einfach: Im Test werden eine Reihe von Fragen zur Ausprägung eines bestimmten Persönlichkeitsmerkmals gestellt wie z. B. Gewissenhaftigkeit. Eine Person beantwortet diese Fragen, indem sie entweder „Ja“ oder „Nein“ ankreuzt oder das Maß der Zustimmung oder Ablehnung zu einer Aussage angibt. Anschließend werden diese Fragen miteinander korreliert und man erhält eine Aussage über die Zuverlässigkeit bzw. interne Konsistenz des Fragebogens. Je stärker die Antworten korrelieren, desto eher messen die Fragen das Gleiche. Damit ist die interne Konsistenz und so auch die Reliabilität des Tests angegeben. Das gleiche Vorgehen könnte man auch zur Einschätzung von Wörterbüchern verwenden. Doch ist Sprache komplexer als die Antwort auf eine Frage in einem Persönlichkeitstest. Wenn eine Person etwas sagt, wird sie es nicht im nächsten oder den darauf folgenden Sätzen wieder sagen, denn normalerweise wird etwas gesagt und dann



zum nächsten Thema übergegangen. Es wäre seltsam, wenn sich eine Person ständig wiederholt. Aber genau das, nämlich die Wiederholung, ist ein grundlegendes Konstruktionsmittel von Fragebogen-Verfahren: Unterschiedlich formulierte Fragen versuchen das gleiche Konstrukt (z. B. Gewissenhaftigkeit) zu erfassen. Da die Messung von Sprache „im Feld“ erfolgt - also eine natürliche Verhaltensprobe darstellt - ist dieses Konstruktionsprinzip für Textanalysen nicht anwendbar. Aufgrund dieser Tatsache haben Textanalyse-Verfahren die Schwierigkeit, mit „komplexen Antworten“ umzugehen.

Oftmals wird eine hohe interne Konsistenz als Voraussetzung für Validität angesehen. Verschiedene Wissenschaftler (z. B. Reuman, 1984) haben darauf hingewiesen, dass – zumindest in der Messung von Motiven in Sprache – geringe interne Konsistenz keinen Einfluss auf die Konstruktvalidität aufweist. Atkinson (1981) verwies darauf, dass die Annahmen der klassischen Testtheorie (nach Gulliksen, 1950) für die Messung von Motiven in Sprache unangemessen seien. Aufgrund der geschilderten Besonderheiten hält Pennebaker (2015) fest, dass die Zuverlässigkeitskoeffizienten der natürlichen Sprache niedriger sind als die, die sonst bei psychologischen Tests gemessen werden.

Aber wie genau sieht das Vorgehen zur Messung und Berechnung der internen Konsistenz der einzelnen Wörterbücher, die 100 Worte verwendet, aus? Die 100 Worte Wörterbücher sind nach psychologischen Merkmalen gruppiert. So gibt es ein Wörterbuch für die Emotion Ärger, eines für das Motiv Leistung, usw. Wir nehmen an, dass mehrere Worte eines Wörterbuchs verwendet werden, wenn über das übergeordnete psychologische Merkmal - bewusst oder unbewusst - gesprochen wird. Wenn eine Person z. B. traurig ist, wird sie verschiedene Worte verwenden, die Trauer ausdrücken. Diese Worte haben also eine höhere Wahrscheinlichkeit, gemeinsam aufzutauchen, als Worte, die zu einem ganz anderen Merkmal gehören (z. B. Freude). Diese gemeinsame Auftretenswahrscheinlichkeit kann gemessen werden und gibt die interne Konsistenz unserer Wörterbücher an.

Abschließend kann zum Thema Reliabilität von Textanalyse-Verfahren Folgendes festgehalten werden: Fragebogen-Verfahren erreichen höhere interne Konsistenz-Werte. Diese führen aber nicht zwangsläufig zu validieren Aussagen. So zeigten sich textbasierte Verfahren zur Messung von Motiven in der Vorhersage von Berufserfolgen prädiktiv valide – sogar 16 Jahre nach Messung. Allerdings zeigten sich im selben Zeitraum keine

Zusammenhänge mit fragebogen-basierten Motivmessungen (z. B. McClelland & Boyatzis, 1982, Jenkins, 1994 oder Collins, Hanges & Locke, 2004). Erklärt werden kann dieser scheinbar widersprüchliche Zusammenhang zwischen mittelmäßiger Reliabilität (=Homogenität) und hoher prädiktiver Validität von Textverfahren durch die Unmittelbarkeit von Sprache. Sie stellt eine unverfälschte, direkte Verhaltensprobe einer Person dar. Fragebögen erfordern stattdessen eine bewusste Auseinandersetzung, da sie kein natürliches Verhalten erfassen. Dadurch sind per se latente (unbewusste) Inhalte nicht zugänglich und durch Vorstellungen davon (=Introspektion) oder durch soziale Erwünschtheit verfälscht. Gerade latente Variablen sind aber langfristig wirksam, wie die zitierten Langzeitstudien zeigen, und deren Berücksichtigung ist unverzichtbar in der Vorhersage von Verhalten.

### Berechnung interner Konsistenz

Zur Prüfung der internen Konsistenz zogen wir Datensets heran, die möglichst verschiedene Sprachanlässe boten, um eine generelle Aussage über die interne Konsistenz treffen zu können. Dieses Vorgehen ist Standard in der Bewertung der Zuverlässigkeit und findet sich auch in anderen Gütemessungen (z. B. in der Bewertung der Internen Konsistenz des Language Inquiry and Word Count, LIWC, Pennebaker, 2008). Dieses Vorgehen empfiehlt sich auch aus der Beobachtung heraus, dass die Verbindung zwischen Sprache und Persönlichkeit zum Teil vom Kontext oder Sprachanlass abhängt, wie Befunde von Mehl und Robbins (2012) nahelegen und unsere Beobachtungen bestätigen. Welche Datensets für die Berechnung der internen Konsistenz verwendet wurden, ist Tabelle 3 zu entnehmen.

Tabelle 3. Datensets zur Berechnung der psychometrischen Eigenschaften

	Formale Sprache			Natürliche Sprache		
	Parlamentsreden	Vorstandsreden	Ausschreibungen	Produktbewertungen	Expressive Writing	Twitter
Beschreibung	Transkriptionen von Reden des europäischen Parlaments	Vorstandsreden zur Aktionärsversammlung	Texte aus Stellenausschreibungen	Produktbewertungen von Amazon Kunden	Geschichten aus der Picture-Story-Exercise (PSE)	Gesammelte Tweets
# Wörter	6.114.599	10.175.519	20.210	5.647.850	107.637	193.039

Anmerkung. Texte, die weniger als 100 Worte enthielten, wurden von der Analyse ausgeschlossen.

Die internen Konsistenzen bewegen sich in einem Feld von 0.22 (Kategorie: Negation) und 0.94 (Kategorie: positive Emotion) mit einem durchschnittlichen Wert von 0.62. Damit

bewegen wir uns etwa im dem Bereich, der auch für LIWC berichtet wird. Eine Auflistung der internen Konsistenzen für die einzelnen Kategorien, findet sich in Tabelle 4.

Tabelle 4. Mittelwerte und Interne Konsistenzen

	Formale Sprache			Natürliche Sprache			Gesamt-Mittelwert	Interne Konsistenz
	Parlamentsreden	Vorstandsreden	Ausschreibungen	Produktbewertungen	Expressive Writing	Twitter		
<b>Personalpronomen</b>	0,049	0,061	0,067	0,04	0,083	0,041	0,341	0.61
<b>Erste Person singular</b>	0,015	0,002	0	0,024	0,006	0,014	0,061	0.51
<b>Erste Person plural</b>	0,016	0,051	0,027	0,001	0,002	0,010	0,107	0.58
<b>Dritte Person singular</b>	0,013	0,007	0,021	0,012	0,061	0,008	0,122	0.38
<b>Artikel</b>	0,156	0,107	0,072	0,099	0,123	0,073	0,63	0.53
<b>Vergangenheit</b>	0,014	0,014	0,003	0,020	0,027	0,014	0,092	0.64
<b>Zukunft</b>	0,024	0,019	0,009	0,011	0,019	0,013	0,095	0.43
<b>Präposition</b>	0,124	0,119	0,124	0,091	0,124	0,101	0,683	0.47
<b>Adverbien</b>	0,111	0,085	0,051	0,147	0,142	0,117	0,653	0.72
<b>Hilfsverben</b>	0,057	0,037	0,018	0,055	0,069	0,042	0,278	0.45
<b>Negationen</b>	0,011	0,004	0,001	0,015	0,012	0,008	0,051	0.22
<b>Schimpfworte</b>	0,003	0,001	0,002	0,007	0,004	0,006	0,023	0.29
<b>Positive Emotion</b>	0,036	0,039	0,057	0,043	0,050	0,036	0,261	0.69
<b>Negative Emotion</b>	0,023	0,005	0,002	0,022	0,018	0,020	0,09	0.65
<b>Angst</b>	0,003	0,001	0,000	0,002	0,003	0,001	0,01	0.33
<b>Ärger</b>	0,004	0,000	0,000	0,004	0,004	0,006	0,018	0.33
<b>Trauer</b>	0,003	0,000	0,000	0,004	0,002	0,002	0,011	0.34
<b>Diskrepanz</b>	0,023	0,011	0,007	0,032	0,033	0,026	0,132	0.53
<b>Absolutheit</b>	0,033	0,021	0,014	0,038	0,028	0,022	0,156	0.54
<b>Machtmotiv</b>	0,027	0,015	0,020	0,011	0,020	0,006	0,099	0.61
<b>Leistungsmotiv</b>	0,025	0,037	0,042	0,014	0,029	0,012	0,159	0.69
<b>Beziehungsmotiv</b>	0,026	0,040	0,041	0,012	0,038	0,021	0,178	0.53
<b>Gewinn</b>	0,021	0,036	0,023	0,014	0,019	0,012	0,125	0.61
<b>Verlust</b>	0,011	0,005	0,004	0,006	0,008	0,006	0,04	0.43
<b>Maskulin</b>	0,009	0,011	0,016	0,004	0,006	0,003	0,049	0.41
<b>Feminin</b>	0,007	0,006	0,010	0,005	0,010	0,006	0,044	0.36

	Formale Sprache			Natürliche Sprache			Gesamt- Mittelwert	Interne Konsistenz
	Parlaments- reden	Vorstands- reden	Ausschreibung- en	Produktbe- wertungen	Expressive Writing	Twitter		
<b>Descriptive Action Verbs</b>	0,027	0,022	0,016	0,024	0,040	0,024	0,153	0.61
<b>State Verbs</b>	0,059	0,040	0,025	0,057	0,071	0,045	0,297	0.55

# LITERATUR

- Atkinson, J. W. (1981). Studying personality in the context of an advanced motivational psychology. *American Psychologist*, 36, 171-128.
- Collins, C. J., Hanges, P. J., & Locke, E. A. (2004). The Relationship of Achievement Motivation to Entrepreneurial Behavior: A Meta-Analysis. *Human Performance*, 17(1), 95-117.
- Gulliksen, H. (1950). Theory of mental tests. New York: Wiley.
- Jenkins, S. R. (1994). Need for power and women's careers over 14 years: Structural power, job satisfaction, and motive change. *Journal of Personality and Social Psychology*, 66(1), 155-165.
- Mackenzie, T. B., Rosenberg, S. D., Bergen, B. J., & Tucker, G. J. (1978). The manipulative patient: an interactional approach. *Psychiatry*, 41(3), 264-271.
- McClelland, D. C., & Boyatzis, R. E. (1982). Leadership motive pattern and long-term success in management. *Journal of Applied Psychology*, 67(6), 737-743.
- Meier, T., Boyd, R. L., Pennebaker, J. W., Mehl, M. R., Martin, M., Wolf, M., & Horn, A. B. (2019). "LIWC auf Deutsch": The Development, Psychometrics, and Introduction of DE-LIWC2015.
- Pennebaker, J. W., & Beall, S. K. (1986). Confronting a traumatic event: toward an understanding of inhibition and disease. *Journal of abnormal psychology*, 95(3), 274.
- Pennebaker, J. W. (2011). The secret life of pronouns. *New Scientist*, 211(2828), 42-45.
- Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*.
- Rorschach, H. (1921). Psychodiagnostik: Methodik und ergebnisse eines wahrnehmungsdiagnostischen Experiments (deutenlassen von zufallsformen) (Vol. 2). E. Bircher.
- Schultheiss, O. C., Riebel, K., and Jones, N. M. (2009). Activity inhibition: a predictor of lateralized brain function during stress. *Neuropsychology* 23, 392–404.

- Schwartz, H. A., Eichstaedt, J., Blanco, E., Dziurzyński, L., Kern, M. L., Ramones, S., ... & Ungar, L. (2013, June). Choosing the right words: Characterizing and reducing error of the word count approach. In *Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 1: Proceedings of the Main Conference and the Shared Task: Semantic Textual Similarity* (pp. 296-305).
  - Tausczik, Y.R., & Pennebaker, J.W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*, 24–54.
  - Veroff, J., Reuman, D., u. Feld, S. (1984). Motives in American men and women across the adult life span. *Developmental Psychology, 20*, 1142-1158.
  - Schiller, A., Teufel, S., Stöckert, C., & Thielen, C. (1999). Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset). *Universität Stuttgart: Institut für maschinelle Sprachverarbeitung*. <http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf>.
2. Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin, 29*, 665-675.
  3. Lepore, S. J. (1997). Expressive writing moderates the relation between intrusive thoughts and depressive symptoms. *Journal of Personality and Social Psychology, 73*, 1030-1037.
  4. Schwartz, H. A., Sap, M., Kern, M. L., Eichstaedt, J. C., Kapelner, A., Agrawal, M., & Kosinski, M. (2016). Predicting individual well-being through the language of social media. In *Biocomputing 2016: Proceedings of the Pacific Symposium* (pp. 516-527).
  5. Zimmermann, J., Brockmeyer, T., Hunn, M., Schauenburg, H., & Wolf, M. (2017). First-person pronoun use in spoken language as a predictor of future depressive symptoms: Preliminary evidence from a clinical sample of depressed patients. *Clinical psychology & psychotherapy, 24*(2), 384-391.
  6. Al-Mosaiwi, M., & Johnstone, T. (2018). In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science, 6*(4), 529-542.
  7. Sonnenschein, A. R. (2016). Anwendung quantitativer Sprachanalyse in der Psychotherapieforschung.
  8. Baddeley, J. L., & Singer, J. A. (2008). Telling losses: Personality correlates and functions of bereavement narratives. *Journal of Research in Personality, 42*(2), 421-438.

9. Block-Lerner, J., Adair, C., Plumb, J. C., Rhatigan, D. L., & Orsillo, S. M. (2007). The case for mindfulness-based approaches in the cultivation of empathy: Does nonjudgmental, present-moment awareness increase capacity for perspective-taking and empathic concern? *Journal of Marital & Family Therapy*, 33, 501-516.
10. Boals, A., & Klein, K. (2005). Word use in emotional narratives about failed romantic relationships and subsequent mental health. *Journal of Language and Social Psychology*, 24(3), 252-268.
11. Burke, P. A., & Dollinger, S. J. (2005). "A Picture's Worth a Thousand Words": Language Use in the Autophotographic Essay. *Personality and Social Psychology Bulletin*, 31(4), 536-548.
12. Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of personality and social psychology*, 90(5), 862.
13. Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of personality and social psychology*, 84(4), 857.
14. Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, 45(3), 211-236.
15. Pennebaker, J. W., Mayne, T. J., & Francis, M. E. (1997). Linguistic predictors of adaptive bereavement. *Journal of personality and social psychology*, 72(4), 863.
16. Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: language use over the life span. *Journal of personality and social psychology*, 85(2), 291.
17. Rude, S., Gortner, E. M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8), 1121-1133.
18. Slatcher, R. B., Vazire, S., & Pennebaker, J. W. (2008). Am "I" more important than "we"? Couples' word use in instant messages. *Personal Relationships*, 15(4), 407-424.
19. Sloan, D. M. (2005). It's all about me: Self-focused attention and depressed mood. *Cognitive Therapy and Research*, 29(3), 279-288.
20. Burke, P. A., & Dollinger, S. J. (2005). "A Picture's Worth a Thousand Words": Language Use in the Autophotographic Essay. *Personality and Social Psychology Bulletin*, 31(4), 536-548.

21. Sexton, J. B., & Helmreich, R. L. (2000). Analyzing cockpit communications: the links between language, performance, error, and workload. *Human Performance in Extreme Environments*, 5(1), 63-68.
22. Pennebaker, J. W. (2011). The secret life of pronouns. *New Scientist*, 211(2828), 42-45.
23. Seider, B. H., Hirschberger, G., Nelson, K. L., & Levenson, R. W. (2009). We can work it out: Age differences in relational pronouns, physiology, and behavior in marital conflict. *Psychology and aging*, 24(3), 604.
24. Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of personality and social psychology*, 77(6), 1296.
25. Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin*, 29(5), 665-675.
26. Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: language use over the life span. *Journal of personality and social psychology*, 85(2), 291.
27. Pasupathi, M. (2007). Telling and the remembered self: Linguistic differences in memories for previously disclosed and previously undisclosed events. *Memory*, 15(3), 258-270.
28. Pennebaker, J. W., & Lay, T. C. (2002). Language use and personality during crises: Analyses of Mayor Rudolph Giuliani's press conferences. *Journal of Research in Personality*, 36(3), 271-282.
29. Dino, A., Reysen, S., & Branscombe, N. R. (2009). Online interactions between group members who differ in status. *Journal of Language and Social Psychology*, 28(1), 85-93.
30. Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, 33(2), 125-143.
31. Robinson, R. L., Navea, R., & Ickes, W. (2013). Predicting final course performance from students' written self-introductions: A LIWC analysis. *Journal of Language and Social Psychology*, 32(4), 469-479.
32. Hancock, J. T., Landrigan, C., & Silver, C. (2007, April). Expressing emotion in text-based communication. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 929-932). ACM.



- 33.Hancock, J. T., Woodworth, M., & Boochever, R. (2018). Psychopaths Online: The Linguistic Traces of Psychopathy in Email, Text Messaging and Facebook. *Media and Communication*, 6(3), 83-92.
- 34.Handelman, L. D., & Lester, D. (2007). The content of suicide notes from attempters and completers. *Crisis*, 28(2), 102-104.
- 35.Bantum, E. O. C., & Owen, J. E. (2009). Evaluating the validity of computerized content analysis programs for identification of emotional expression in cancer narratives. *Psychological assessment*, 21(1), 79.
- 36.D'Souza, P. J., Lumley, M. A., Kraft, C. A., & Dooley, J. A. (2008). Relaxation training and written emotional disclosure for tension or migraine headaches: a randomized, controlled trial. *Annals of Behavioral Medicine*, 36(1), 21-32.
- 37.Gillis, M. E., Lumley, M. A., Mosley-Williams, A., Leisen, J. C., & Roehrs, T. (2006). The health effects of at-home written emotional disclosure in fibromyalgia: A randomized trial. *Annals of Behavioral Medicine*, 32(2), 135-146.
- 38.Gaucher, D., Friesen, J., & Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of personality and social psychology*, 101(1), 109.
- 39.Konrad, A. M., Ritchie Jr, J. E., Lieb, P., & Corrigan, E. (2000). Sex differences and similarities in job attribute preferences: a meta-analysis. *Psychological bulletin*, 126(4), 593.
- 40.Cassell, J., Huffaker, D., Tversky, D., & Ferriman, K. (2006). The language of online leadership: Gender and youth engagement on the Internet. *Developmental psychology*, 42(3), 436.
- 41.Winter, D. G. (1991). A motivational model of leadership: Predicting long-term management success from TAT measures of power motivation and responsibility. *The Leadership Quarterly*, 2(2), 67-80.
- 42.Schultheiss, O. C. (2013). Are implicit motives revealed in mere words? Testing the marker-word hypothesis with computer-based text analysis. *Frontiers in psychology*, 4, 748.
- 43.Smith, C. P., Atkinson, J. W., McClelland, D. C., & Veroff, J. (Eds.). (1992). *Motivation and personality: Handbook of thematic content analysis*. Cambridge University Press.
- 44.Danescu-Niculescu-Mizil, C., Lee, L., Pang, B., & Kleinberg, J. (2012, April). Echoes of power: Language effects and power differences in social interaction. In *Proceedings of the 21st international conference on World Wide Web* (pp. 699-708). ACM.

45. Steinmann, B., Dörr, S. L., Schultheiss, O. C., & Maier, G. W. (2015). Implicit motives and leadership performance revisited: What constitutes the leadership motive pattern?. *Motivation and Emotion*, 39(2), 167-174.
46. Jenkins, S. R. (1994). Need for power and women's careers over 14 years: Structural power, job satisfaction, and motive change. *Journal of Personality and Social Psychology*, 66(1), 155.
47. Johannßen, D., Biemann, C., & Scheffer, D. (2019, June). Reviving a psychometric measure: Classification and prediction of the Operant Motive Test. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology* (pp. 121-125).
48. Collins, C. J., Hanges, P. J., & Locke, E. A. (2004). The relationship of achievement motivation to entrepreneurial behavior: A meta-analysis. *Human performance*, 17(1), 95-117.
49. Köhler, A. (2009). Zur Validität reaktionszeitbasierter Motivmessung im Kontext der Personalauswahl.
50. Lee, A. Y., Keller, P. A., & Sternthal, B. (2009). Value from regulatory construal fit: The persuasive impact of fit between consumer goals and message concreteness. *Journal of Consumer Research*, 36(5), 735-747.
51. Higgins, E. T. (2005). Value from regulatory fit. *Current directions in psychological science*, 14(4), 209-213.
52. Lee, A. Y., & Aaker, J. L. (2004). Bringing the frame into focus: the influence of regulatory fit on processing fluency and persuasion. *Journal of personality and social psychology*, 86(2), 205.
53. Semin, G. R., Higgins, T., de Montes, L. G., Estourget, Y., & Valencia, J. F. (2005). Linguistic signatures of regulatory focus: how abstraction fits promotion more than prevention. *Journal of personality and social psychology*, 89(1), 36.
54. Collier, D., & Zhang, C. (2016). Can We Reduce Bias in the Recruiting Process and Diversify Pools of Candidates by Using Different Types of Words in Job Descriptions?.
55. [https://www.100worte.de/wp-content/uploads/2019/02/Abstract\\_FunkeBarth.pdf](https://www.100worte.de/wp-content/uploads/2019/02/Abstract_FunkeBarth.pdf)
56. [https://www.100worte.de/wp-content/uploads/2019/07/100Worte\\_Validierung\\_Motive.pdf](https://www.100worte.de/wp-content/uploads/2019/07/100Worte_Validierung_Motive.pdf)
57. [https://www.100worte.de/wp-content/uploads/2019/02/Artikel\\_Genderbias.pdf](https://www.100worte.de/wp-content/uploads/2019/02/Artikel_Genderbias.pdf)

- 58.Hartley, J. (2003). Improving the clarity of journal abstracts in psychology: the case for structure. *Science Communication*, 24(3), 366-379.
- 59.Burke, P. A., & Dollinger, S. J. (2005). "A Picture's Worth a Thousand Words": Language Use in the Autophotographic Essay. *Personality and Social Psychology Bulletin*, 31(4), 536-548.
- 60.Centerbar, D. B., Schnall, S., Clore, G. L., & Garvin, E. D. (2008). Affective incoherence: When affective concepts and embodied reactions clash. *Journal of personality and social psychology*, 94(4), 560.
- 61.Rogozinska, D., & Wawer, A. (2013, December). Interpreting or describing? Measuring verb abstraction. In *2013 IEEE 13th International Conference on Data Mining Workshops* (pp. 963-966). IEEE.
- 62.Holmes, D., Alpers, G. W., Ismailji, T., Classen, C., Wales, T., Cheasty, V., ... & Koopman, C. (2007). Cognitive and emotional processing in narratives of women abused by intimate partners. *Violence against women*, 13(11), 1192-1205.

# ANHANG

## BEGRIFFSERKLÄRUNG

<sup>1</sup> Bei der statistischen Analyse der binären Klassifizierung ist der F1-Score (auch F-Score oder F-Messung) ein Maß für die Genauigkeit eines Tests. Es berücksichtigt sowohl die Genauigkeit  $p$  als auch die Trefferquote  $r$  des Tests:  $p$  ist die Anzahl der korrekten positiven Ergebnisse dividiert durch die Anzahl aller positiven Ergebnisse, die vom Klassifikator zurückgegeben werden,  $r$  ist die Anzahl der korrekten positiven Ergebnisse dividiert durch die Anzahl aller relevanten Proben (alle Proben, die als positiv identifiziert werden sollten). Der F1-Wert ist der harmonische Mittelwert aus Genauigkeit und Trefferquote, wobei ein F1-Wert seinen besten Wert bei 1 (perfekte Genauigkeit und Trefferquote) und seinen schlechtesten bei 0 erreicht.

<sup>2</sup> Statistisch signifikant wird das Ergebnis eines statistischen Tests genannt, wenn Stichprobendaten so stark von einer vorher festgelegten Annahme (der Nullhypothese) abweichen, dass diese Annahme nach einer vorher festgelegten Regel verworfen wird. Sinnvollerweise wird bei der Festlegung dieser kritischen Schwelle bedacht, welche Konsequenzen der Fall hätte, dass irrtümlich angenommen wird, ein beobachteter Unterschied sei nur zufällig. Hält man diese Folgen eher für gravierend, so wird man hier eher ein niedriges Niveau als ein höheres wählen, beispielsweise lieber 1 % als 5 %, oder aber 0,1 % für die maximal zulässige Irrtumswahrscheinlichkeit festlegen. Diese Wahrscheinlichkeit wird als Signifikanzniveau  $\alpha$  bezeichnet. So bedeutet  $\alpha=0,05$ : Falls die Nullhypothese richtig ist, darf die Wahrscheinlichkeit dafür, dass sie fälschlich abgelehnt wird (Fehler 1. Art), nicht mehr als 5 % betragen. Entsprechend beträgt dann die Wahrscheinlichkeit, eine richtige Nullhypothese aufgrund des statistischen Tests nicht abzulehnen,  $1-\alpha=0,95$ , sprich mindestens 95%.

<sup>3</sup> Jede wissenschaftliche Messmethode muss bestimmten Gütekriterien genügen. Objektivität und Zuverlässigkeit sind zentrale Forderungen an eine Messung. Als Hauptgütekriterien gelten:

- *Objektivität*: Sind die Ergebnisse unabhängig von Einflüssen der Untersucher oder der Untersuchungssituation bei Durchführung, Auswertung und Interpretation zustande gekommen?
- *Reliabilität*: Wird das Merkmal zuverlässig gemessen oder ist die Messung in zu großem Ausmaß mit Messfehlern behaftet?
- *Validität*: Misst das Verfahren tatsächlich das gewünschte Merkmal? Ist die Verwendbarkeit des Verfahrens für eine diagnostische Entscheidung gegeben?

Quelle: Wikipedia