Krisztián Herbert J.

Überzufälligkeit, Musterhaftigkeit und Persönlichkeitsspezifizität im menschlichen Sprachgebrauch

Sprachstatistik auf den Spuren des sprachlichen Fingerabdrucks*

Die vorliegende Arbeit verwendet sprachstatistische und korpuslinguistische Methoden unter Einbeziehung von 16 sprachstatistischen Indikatoren, angewendet auf ein Gesamtkorpus von 10.000 analysierten Texteinheiten aus der deutschen Online-Blog-Schriftsprache, verfasst von insgesamt 55 Sprachbenutzern. Die statistische Analyse wurde computergestützt mit Python durchgeführt, um die jeweiligen Kohärenzmuster des persönlichkeitsspezifischen Sprachgebrauchs zu identifizieren. Die empirische Untersuchung konzentriert sich auf zwei Hypothesen, die sich auf die Kohärenz bzw. Diskrepanz der Indikatorwerte beziehen, wodurch eine automatische Erstellung von Persönlichkeitsprofilen ermöglicht wurde.

Schlüsselwörter:
Psycholinguistik, kognitive Linguistik, Überzufälligkeit, sprachstatistische Indikatoren, sprachlicher Fingerabdruck, persönlichkeitsspezifische Manifestationen, Python

1. Einführung

Die vorliegende Arbeit ist eine Untersuchung im Sinne einer ‚Research' innerhalb der germanistischen Psycholinguistik. Ziel dieser Arbeit ist es, persönlichkeitsspezifische Sprachgebrauchsmuster in sprachlichen Strukturen auf der Sprachoberfläche zu ermitteln. Die Arbeit geht davon aus, dass die daktyloskopischen Spuren der menschlichen Persönlichkeit im menschlichen Sprachgebrauch explizit erscheinen (vgl. Schubert 2008: 5), da die kognitiven Informationsverarbeitungsmodi des menschlichen Persönlichkeits- und Sprachverarbeitungssystems unvermeidlich den Wahrnehmungsrahmen des einzelnen Individuums bilden können (vgl. Schubert 2008: 5). „In jedem winzigen Wort ist jedoch auch nicht nur die semantische Geschichte der Sprachgemeinschaft enthalten, vielmehr auch die häufig nicht bewusste persönliche Biografie des Menschen", so Herbert Bock (2021).

Chomsky (2010: 116) geht davon aus, dass das Individuum über ein bestimmtes, kognitives Wissenssystem verfügt, das sowohl im menschlichen Geist als auch in den neuronalen Konfigurationen des menschlichen Gehirns repräsentiert sein muss. Es stellt sich die Frage, wie ein solches Wissenssystem überhaupt vorstellbar ist oder was die neuronalen Netzwerke eines

* Eingereicht wurde die vorliegende, verkürzte Arbeit als Teil der Masterarbeit „Persönlichkeit und Sprachvariation. Persönlichkeitsspezifische Manifestationen in sprachlichen Strukturen" im Frühlingssemester 2025. Die ursprüngliche Arbeit entstand während eines Auslandsemesters mit dem Schwerpunkt Allgemeine Linguistik an der Ruprecht-Karls-Universität Heidelberg. Die Arbeit wurde von Robert Rada betreut. Erreichbarkeit des Autors: herbert.j.krisztian@gmail.com.

Sprachbenutzers genau widerspiegeln würden. Er zieht die Konsequenz, dass diese Fragen „in den Bereich der Linguistik und Psychologie“ fallen, „zwei Gebiete, die ich eigentlich gar nicht trennen möchte“ (2010: 116). Die Aufmerksamkeit des Forschers wird dabei mit einem höchstkomplexen Wahrnehmungs- und Produktionsproblem konfrontiert. Chomsky sieht in der Linguistik ein Potenzial für eine Wissenschaft, die mit naturwissenschaftlichem Anspruch die Natur des Menschen erforscht:

> Soweit der Linguist Antworten auf die Fragen […] geben kann, kann der Neurowissenschaftler beginnen, die physischen Mechanismen zu erforschen, die die von der abstrakten Theorie des Linguisten enthüllten Eigenschaften aufweisen. Solange es keine Antworten auf diese Fragen gibt, wissen die Neurowissenschaftler nicht, wonach sie suchen; ihre Forschung ist dann in dieser Hinsicht blind. (Chomsky 2010: 117)

In den letzten Jahren gab es zahlreiche Neuerungen im Bereich der großen sprachlichen Modelle (vgl. Pichler/Pramer 2023). Diese Modelle bergen ein vielversprechendes Potenzial für eine neue Ära der wissenschaftlichen Methodologie – sowohl in den theoretischen als auch in den angewandten Wissenschaften. Sie revolutionieren u.a. die maschinelle Sprachverarbeitung (vgl. Zenthöfer 2023) und können zukünftige sprachdiagnostische Instrumente ermöglichen (vgl. Hoffmann 2023). Damit ließe sich die seit der kognitiven Wende bestehende Herausforderung bewältigen, einen tieferen Einblick in die Black Box der Mechanismen der menschlichen Wahrnehmung zu gewinnen. Lakoff und Wehling reflektieren diese Aufgabe wie folgt:

> Und ich habe auch eine ausgesprochen realistische Auffassung davon, wie wir Menschen die Welt denkend begreifen: Menschen begreifen die Welt nicht durch irgendeine abstrakte universelle Instanz, die wir „Verstand“ nennen und die jenseits unserer Körperlichkeit liegt. Menschen begreifen die Welt durch ihr Gehirn, das ein Teil des Körpers ist. Sie denken auf Grund von Konzepten, die physisch in ihrem Gehirn vorhanden sind. Denken ist physisch. (Lakoff/Wehling 2010: 151)

Nach dieser Interpretation besteht eine der signifikanten Aufgaben der Psycholinguistik heute darin, die Korrelationen zwischen Sprachgebrauch und Persönlichkeit zu entschlüsseln, um die Navigation zwischen den Zusammenhängen zu ermöglichen. Dementsprechend beabsichtigt die vorliegende Arbeit, einen Versuch zu unternehmen, die persönlichkeitsspezifischen Sprachgebrauchsmuster durch korpuslinguistische und statistische Methoden sichtbar zu machen. Allerdings sind die Grenzen einer textlinguistischen Korpusanalyse der deutschen Schriftsprache schon deutlich. Wie der Vater des kognitiven Projekts, Noam Chomsky, formulierte: „The trouble is, there has not in history ever been any answer other than, ‚Get to work on it‘“ (Chomsky 1995). Die vorliegende Arbeit verwendet sprachstatistische und korpuslinguistische Methoden mit der Einbeziehung von 16 sprachstatistischen Indikatoren an

einem Gesamtkorpus von 10.000 analysierten Texteinheiten der deutschen Online-Blog-Schriftsprache von insgesamt 55 Sprachbenutzern. Die statistische Analyse ist computergestützt mit Python durchgeführt worden, um die jeweiligen Kohärenzmuster des persönlichkeitsspezifischen Sprachgebrauchs zu ermitteln. Die empirische Arbeit wurde auf die Untersuchung von zwei Hypothesen begrenzt. Die einzelnen statistischen Matrizen, die die signifikanten Übereinstimmungen und Abweichungen belegen, beinhalten schon beim Vergleich der Werte der fünf Personen 720 quantifizierte Angaben pro Tabelle. Die untersuchten Hypothesen sind die folgenden:

Die erste Hypothese zur Wertekonstanz des Sprachgebrauchsmusters eines Sprachbenutzers geht davon aus, dass Persönlichkeit und Sprachgebrauch eng zusammenhängen, da sie auf einem gemeinsamen neuronalen Hintergrund basieren. Dieser Zusammenhang führt zu überzufälligen Sprachgebrauchsmustern, in denen kognitive Wahrnehmungsmuster eine relativ stabile statistische Wertekonstanz sprachlicher Strukturen auf der Sprachoberfläche bewirken. Wenn die Situativität des Sprachgebrauchs mithilfe korpuslinguistischer Methoden weitgehend ausgeschaltet wird, sollten die 16 sprachstatistischen Indikatoren kohärente Sprachgebrauchsmuster des jeweiligen Sprachbenutzers aufzeigen. Daraus folgt: Vergleicht man die Indikatorwerte von 1.000 zufällig ausgewählten Texteinheiten desselben Verfassers – aufgeteilt im Verhältnis 500:500 –, so sollten sich diese als statistisches Spiegelbild der ermittelten Werte erweisen. Bei einem Signifikanzniveau von 0,01 wären demnach keine signifikanten Unterschiede zu erwarten.

Die zweite Hypothese zur Diskrepanz zwischen den Sprachgebrauchsmustern verschiedener Sprachbenutzer basiert auf der Annahme, dass Persönlichkeit und Sprachgebrauch eng miteinander verknüpft sind, da sie auf einem gemeinsamen neuronalen Hintergrund basieren. Dadurch entstehen überzufällige Sprachgebrauchsmuster, deren statistische Wertetendenzen sich auf der Sprachoberfläche manifestieren. Wenn diese Muster keinem Gruppierbarkeitskriterium unterliegen, lassen sie sich voneinander abgrenzen. Wird die Situativität des Sprachgebrauchs mithilfe korpuslinguistischer Methoden weitgehend ausgeschlossen, sollten die 16 sprachstatistischen Indikatoren die voneinander unterscheidbaren Sprachgebrauchsmuster aller Sprachbenutzer sichtbar machen. Konkret bedeutet dies: Vergleicht man die Indikatorwerte von jeweils 1.000 zufällig ausgewählten Texteinheiten eines Verfassers – aufgeteilt im Verhältnis 500:500 –, so sollten sie ein statistisches Spiegelbild der ermittelten Werte ergeben, das bei einem Signifikanzniveau von 0,01 keine signifikanten Unterschiede aufweist. Anders verhält es sich jedoch, wenn die Indikatorwerte der im Verhältnis 500:500 verteilten Texteinheiten zwischen verschiedenen

Sprachbenutzern verglichen werden. In diesem Fall sollten sie sich in mehreren Indikatorwerten signifikant voneinander unterscheiden ($p < 0,01$), sofern kein Gruppierbarkeitskriterium vorliegt. Ein solches Kriterium wäre nur dann gegeben, wenn sich die Indikatorwerte unterschiedlicher Sprachbenutzer bei einem Signifikanzniveau von 0,01 nicht signifikant unterscheiden.

2. Die theoretischen Grundlagen der Persönlichkeits-Interaktions-Theorie bei der Untersuchung des differentiellen Ausdrucksverhaltens als Rahmenkonzept

Die vorliegende Arbeit verwendet die Kuhl'sche Persönlichkeits-Interaktions-Theorie als kognitiv-emotionalen Deutungsrahmen für die statistischen Ergebnisse (vgl. Kuhl/Quirin/Koole 2020: 7). Nach der PSI-Theorie lässt sich die menschliche Persönlichkeit als ein Set von Merkmalen eines Informationsverarbeitungssystems definieren. Dieses System bestimmt eine relativ stabile kognitive Einstellung der Wahrnehmungsprozesse (vgl. Kuhl/Quirin/Koole 2020: 7). Persönlichkeitsunterschiede beruhen demnach auf charakteristischen Systemkonfigurationen, die jeweils eine spezifische Informationsverarbeitungsform prägen. Die schematisierte Matrix der einzelnen Systeme bzw. Systemfunktionen ergibt das folgende Gesamtbild:

hoch-inferente Systeme	
Intentionsgedächtnis (IG)	Extensionsgedächtnis (EG)
sequenziell-analytisch, explizites Wissen (Absichten, Pläne), Entweder-Oder-Charakteristik, Reduktionismus, Emotionsentkopplung: Ich-Bezug, zielfokussierte Aufmerksamkeit, bewusst, Vulnerabilität bei unvollständigen Informationen: Komplexitätsreduktion	parallel-holistisch, implizites Konfigurationswissen: Erwartungen, allgemeine Ziele, Integration von Gegensätzen, Unterscheidungssensitivität, Emotionswahrnehmung, Emotionsregulation, kongruenzbetonte, verteilte Aufmerksamkeit, nicht bewusst, Robustheit be unvollständigen Informationen: Komplexitätstoleranz
Funktion: Denken (AD)	Funktion: Fühlen (GF)
hoch-infenrente, symbolische Repräsentationen, sprachabhängig, hierarchisch, sequenziell, präzise, rigide (keine Lückenresistenz), Emotionskopplung	rational, realitätsbasiert, integriert bewusste Erfahrungen, Gegensätze, höhere Form der Integration, abstrahiert aus Episoden, unterschiedssensitiv, Affinität zur Selbstregulation, Dekomponierbarkeit

niedrig-inferente Systeme	
Intuitive Verhaltenssteuerung (IVS)	Objekterkennung (OES)
parallel-holistisch, intuitive Programme, kontextualisiert, multimodale Verschmelzung, gegenwarts- und zukunftsorientiert, prototypisch, nicht bewusst	sequenziell-analytisch, dekontextualisierten, Separierung verschiedener Sinne, vergangenheitsorientiert („Wiedererkennen), kategorial, bewusst, unstimmigkeitsbetonte Aufmerksamkeit
Funktion: Intuieren (IV)	Funktion: Empfinden (OE)
vom Boden abgekommen, geringe bewusste Repräsentanz, elementares Integrationspotential, lernen durch sensomotorische Rückmeldung, undifferenziert, Affinität zur automatischen Steuerung, Verschmelzung	elementare, symbolische Repräsentation, modalitätssepzifisch, vergangenheitsorientiert (Vergleich mit Vertrauten), allozentrische Wahrnehmung, Kontextabstraktion, Kategorisierungsfokus, motorische Entkopplung

Abb. 1: Funktionsprofile und ausgewählte Merkmale der vier persönlichkeitsrelevanten Makrosysteme, modifiziert nach Kuhl (Schubert 2008: 67, 72, 74)

Nach Schubert lassen sich die kognitiv-emotionalen Verarbeitungsstile der PSI-Theorie mit sprachstatistischen Indikatoren verbinden, die eine Indikatorfunktion für das prädominierende Verarbeitungsmodus erfüllen können (vgl. Schubert 2008: 178). Es lässt sich feststellen, dass ein besseres Verständnis der Zusammenhänge zwischen den sprachstatistischen Indikatoren und den jeweiligen Informationsverarbeitungsstilen noch weiterer Untersuchungen bedarf. Obwohl sich die Vermutungen auf der Sprachoberfläche – wie auch im praktischen Teil gezeigt – als durchaus tragfähig erwiesen haben, wären zur genauen Bestimmung des Geltungsbereichs die jeweiligen neuronalen Korrelate mithilfe bildgebender Verfahren wie EEG und MRT zu erfassen (vgl. Müller 2013: 30). Die sprachstatistischen Indikatoren bzw. deren vermutete Persönlichkeitsspezifizität, die die auf der Sprachoberfläche identifizierbaren Sprachgebrauchsmuster quantifizieren, ergeben folgendes schematisiertes Gesamtbild:

Parameter	Indikator für:	vermutete Korrelation mit analytischen Stilvertretern:	vermutete Korrelation mit ganzheitlichen Stilvertretern
Dogmatismus (DQ)	Geschlossenheit des Denksystems	negativ	positiv
referentielle Prägnanz (D1235)	Tendenz zur Generalisierung	negativ	positiv
operative Prägnanz (D46)	Kohärenz des Denksystems	positiv	negativ
dogmatische-A-Ausdrücke (D(A))	Komplexitätsreduzierung	negativ	positiv
dogmatische-B-Ausdrücke (D(B))	Komplexitätstoleranz	positiv	negativ
Abstraktheit (AI)	Verdichtungen im Denksystem	positiv	negativ
Variationsindex (VAI)	Wortschatz, Informationsmenge	negativ	positiv
Negationen	Vermeidung von Ambiguität	positiv	negativ
begründende Konjunktionen	Verknüpfungsfähigkeit	negativ	positiv
adversative Konjunktionen	Abgrenzung	negativ	positiv
Egozentriertheit	Ich-Nähe, soziale Ordnung	negativ	positiv
Konjunktive	Vorstellungskraft	negativ	positiv
Textlänge	Informationsmenge	negativ	positiv
Subordinationsindex (SUI)	Umgang mit komplexen Satzeinheiten	negativ	positiv
Modalverben	Einbeziehen externer Randbedingungen	negativ	positiv
Hilfsverben	nicht bekannt	nicht bekannt	nicht bekannt

Abb. 2: Vermutete Beziehungen der sprachstatistischen Indikatoren zu den Informationsverarbeitungsstilen (Schubert 2008: 180)

Nach Schwibbe ist es möglich, die „Variation eines Verfassers auf einer dieser Persönlichkeitsdimensionen […] durch die Variation seiner Sprache“ zu erfassen (zit. n. Schubert 2008: 112). Er interpretiert das Ergebnis der sprachanalytischen Indikatoren als Spurelemente der jeweiligen „kognitiven Informationsverarbeitungssysteme“, „die mit verschiedenen Persönlichkeitsdimensionen assoziiert sind“ (Schubert 2008: 112). Das visionierte Endziel einer solchen Analyse wäre es, die persönlichkeitsspezifische Sprachprofilierung des jeweiligen Sprachbenutzers nach der Quantifizierung der

Manifestationen, die auf der Sprachoberfläche erscheinen, zu ermöglichen (vgl. Oakes 2022: 1174). Nach Strasser lässt sich das menschliche Temperament als eine überwiegende Affektlage verstehen, zu der das Zusammenspiel verschiedener psychisch-physischer Teilsysteme beiträgt. Diese dominante Affektlage wird u.a. durch das Persönlichkeits-Stil-und-Störungs-Inventar theoretisiert, dessen Stilebenen laut Schubert durch semantisch-kognitiv orientierte Indikatoren statistisch identifizierbar sind. Dabei ist festzustellen, dass das PSS-Inventar bzw. die PSI-Theorie grundlegende Züge der Jung'schen Psychologie aufweisen (vgl. Jung 2017: 376). Sie entwerfen ein Navigationssystem mit einer positiven Achse – der Ebene der Selbstentfaltung – und einer negativen, pathologischen Achse, die Funktionsstörungen abbildet. Eine visionäre Perspektive verfolgt das Ziel, mithilfe computerlinguistischer Methoden künftig ein Niveau der Quantifizierung zu erreichen, das eine softwaregestützte Navigation zwischen der positiven und der pathologischen Achse mittels sprachdiagnostischer Instrumente ermöglichen könnte (vgl. Oakes 2022: 1174).

Die Verbindung der Sprachproduktionsebenen mit den kognitiven Informations-verarbeitungsmodi, die – nach tiefgehenden Korrelations- und Situationsforschungen – als sprachliche Persönlichkeitsstile gruppiert werden können, bietet bereits einen spannenden Deutungsrahmen für die Analyseergebnisse. In diesen Ergebnissen können die semantisch-kognitiv motivierten sprachstatistischen Indikatoren als kognitive Sprachgebrauchsmuster der individuellen Sprachvariation auf der Sprachoberfläche identifiziert werden. Ein weiterer Interpretationsschritt besteht darin, die jeweiligen sprachlichen Stilgrenzen bzw. vermutete Allostile zu deuten, da die Deutungsrahmen von Kuhl/Kazén mehrere Aktualisierungs-möglichkeiten zulassen. Die von Schubert zur Sprachprofilierung empfohlenen Kategorien der Stilebene des PSS-Inventars werden in der folgenden schematischen Darstellung aufgelistet:

Stil	dominante kognitive Funktionen	bevorzugte Informations-verarbeitung	Stil	dominante kognitive Funktionen	bevorzugte Informations-verarbeitung
selbstbestimmt	Fühlen	ganzheitlich	ehrgeizig	Fühlen, Intuieren	ganzheitlich
eigenwillig	Denken, Fühlen	ausgeglichen	kritisch	Denken, Fühlen	ausgeglichen
zurückhaltend	Denken	analytisch	loyal	Empfinden, Denken	analytisch
selbstkritisch	Denken, Empfinden	analytisch	eigenwillig	Denken, Fühlen	ausgeglichen
sorgfältig	Empfinden	analytisch	spontan	Empfinden, Intuieren	ausgeglichen
ahnungsvoll	Empfinden, Intuieren	ausgeglichen	liebenswürdig	Intuieren	ausgeglichen
optimistisch	Intuieren	ganzheitlich	still	Denken, Empfinden	analytisch
hilfsbereit	Fühlen, Intuieren	ganzheitlich			

Abb. 3: Übersicht der „individuellen Stile“, der vermuteten „kognitiven Dispositionen“ sowie der „bevorzugten Informationsverarbeitungsform“ nach dem PSSI, modifiziert nach Kuhl & Kazén (Schubert 2008: 78)

3. Die persönlichkeitsspezifischen Sprachgebrauchsmuster nach den sprachstatistischen Indikatorwerten

Im Folgenden möchte ich die konkreten 16 sprachstatistischen Indikatoren darstellen, deren Analyseergebnisse die hypothetische Sprachprofilierung des Sprachbenutzers nach den eben erwähnten Informationsverarbeitungsmodi der PSI-Theorie und der Stilebene des PSS-Inventars ermöglichen. Sie bilden die Schlüssel zur Quantifizierung von Zuständen auf der Sprachoberfläche, die die jeweiligen Sprachgebrauchsmuster der individuellen Sprachvariation statistisch zugänglich machen (vgl. Oakes 2022: 1174).

3.1 Abstraktheitsindex als Indikator für die Verdichtungen im Denksystem

Der Abstraktheitsindex kann als Indikator für Verdichtungen im Denksystem betrachtet werden (vgl. Schubert 2008: 178). Die Abstraktheit eines Textes zeigt einen binären Wert zwischen Konkretheit und Abstraktheit. Nach Pikas deutet Schubert (2008: 126) die Abstraktheit als einen „isolierenden“, „generalisierenden“, symbolisierenden und „strukturgenerierenden“ Prozess. Klix sieht das Wesen der Abstraktheit in einer „symbolisierenden Funktion“ als eine „Leistungsdimension abstraktiven Verhaltens“ (Schubert 2008: 127). Nach Schwibbe werden

„abstraktive Verdichtungen“ vorgenommen, „um große Informationsmengen verarbeiten zu können“ (Schubert 2008: 127). „Bei geringer Informationsmenge ist dieser Prozess nicht nötig, da die Einzelelemente noch zu überblicken sind“ (Schubert 2008: 127). Die Abstraktionsleistung ermöglicht die „Entdeckung neuer Handlungswege“ (vgl. Schubert 2008: 325). Die Abstraktionsebene stellt gleichzeitig das „Konstrukt ‚lexikalischer Komplexität‘“ dar (Schubert 2008: 127).

Die Suffixe, die in der Berechnung des Abstraktheitsindex nach Günther und Groeben eine signifikante Rolle spielen, sind die folgenden: -heit (z.B. Abstraktheit), -ie (z.B. Liturgie), -ik (z.B. Grammatik), -ion (z.B. Nation), -ismus (z.B. Konservativismus), -ität (z.B. Identität), -keit (z.B. Häufigkeit), -nz (z.B. Intelligenz), -tur (z.B. Kultur), -ung (z.B. Satzung). Eine Ausnahmengruppe bilden die -ie-Suffixe, wenn sie nicht als langes ‚i‘ ausgesprochen werden. Werden sie wie ‚je‘ (z.B. Familie) ausgesprochen (vgl. Schubert 2008: 200), bleiben sie in der Berechnung außer Acht. Das Abstraktheitskonzept geht davon aus, dass Wörter als binäre Oppositionen entweder als abstrakt oder als konkret berechnet werden können.

Schwibbe interpretiert Dogmatismus und Abstraktheit als „informationsverarbeitende kognitive Systeme“. Während die dogmatische Denkstruktur gewisse Informationen ausschließt, ermöglichen die abstrahierenden Denkstrukturen die Informationsintegration bzw. Verdichtung (zit. n. Schubert 2008: 131). Es gibt Gruppen von Wörtern, die trotz einer abstrakten Endung keinesfalls als abstrakte Größen interpretiert werden können (z.B. Wohnung, Ausbildung), oder sie haben zumindest seit den 1970er-Jahren jene abstrakten Eigenschaften verloren. Diese Wörter gehören organisch zu unserem Alltag und signalisieren keinesfalls eine abstrakte Informationsverarbeitung. Aus diesem Grund werden die erwähnten – ja häufigen – zwei Wörter aus der Berechnung ausgeschlossen. Formel zur Berechnung:

$$Abstraktheitsindex\ (AI) = \frac{\sum S_A}{Textlänge}$$

1. Formel: Die Berechnung des Abstraktheitsindexes (vgl. Schubert 2008: 178)

3.2 Variationsindex als Indikator für Wortschatz und Informationsmenge

Bei Schubert (2008: 178) erscheinen die modifizierten Ergebniswerte des TTR als Indikator für die Variation bzw. Variabilität des Wortschatzes. Da es sich im Falle des TTR bzw. des Variationsindexes bei Schubert eher um eine Überlegung handelt, die der Datenvisualisierung dienen soll, werde ich im Folgenden ausschließlich das TTR als Variationsindex verwenden. Die Verlagerung der TTR-Werte nach den Peak-/Bottom-Werten zwischen 0 und 1 erleichtert die Datenvisualisierung erheblich. Das TTR wurde in der vorliegenden Arbeit unter Eliminierung bestimmter Wortgruppen berechnet. Flexionen, Konjugationen und Komparative,

Präpositionen, Personalpronomen, Reflexivpronomen, Demonstrativpronomen, Hilfsverben, Modalverben, neben- und unterordnende Konjunktionen, Fragewörter sowie bestimmte und unbestimmte Artikel wurden softwaregestützt aus der Berechnung ausgeschlossen. „Der TTR errechnet sich als Quotient aus der Anzahl verschiedener Wörter (Types) und der Gesamtanzahl aller Wörter (Tokens) eines Textes“ (Schubert 2008: 125). Laut Hörmann gilt der TTR-Index als „Indikator für den Wortschatz“ bzw. die lexikalische Variabilität eines Sprachbenutzers (Schubert 2008: 125). Allerdings beeinflussen situative Faktoren, wie beispielsweise die jeweilige Textsorte, die TTR-Werte (Schubert 2008: 125). Schubert verwendet zur Berechnung des TTR die folgende Formel:

$$Type\ Token\ Ratio\ (TTR) = \frac{types}{tokens}$$

2. Formel: Die Berechnung des Type-Token-Ratios (vgl. Schubert 2008: 178)

3.3 Egoquotient als Indikator für Ich-Nähe und soziale Orientierung

Der Egoquotient kann als Indikator für Ich-Nähe und soziale Orientierung verstanden werden (vgl. Schubert 2008: 178). „Dieses Verhaltensmerkmal ist durch ein Verhaftetsein des Verhaltens (und damit auch des Denkens und Sprechens) an die konkrete Situation sowie an die in ihr unmittelbar gemachten Erfahrungen und Wahrnehmungen gekennzeichnet“ (Schubert 2008: 153). Formel zur Berechnung der relativen Häufigkeit des Personalpronomens „Ich“:

$$Egoquotient\ (EQ) = \frac{\sum Perspnalpronomina\ 1.\ Pers.\ Sing.}{Textlänge}$$

3. Formel: Die Berechnung des Egoquotienten (vgl. Schubert 2008: 178).

3.4 Subordinationsindex als Indikator für den Umgang mit komplexen Satzeinheiten und Satzgliederungen

Der Subordinationsindex kann als Indikator für den Umgang mit komplexen Satzeinheiten bzw. für die Satzkomplexität angesehen werden (vgl. Schubert 2008: 178). „Dieser Index bildet sich aus dem Quotienten von Nebensätzen und Hauptsätzen und gibt an, in welche grammatikalische Komplexität die Informationen sprachlich eingebettet sind“ (Schubert 2008: 137). Allerdings zeigt sich die Schwäche der computergestützten Berechnung bereits darin, dass die an der Interpunktion orientierte Berechnungsmethode nur dann funktioniert, wenn der Verfasser über eine fehlerfreie Grammatik verfügt. Lässt der Verfasser die Interpunktionszeichen weg, können sich schwierige Dilemmata ergeben, ob es sich um Haupt- oder Nebensätze handelt. Die vorliegende Arbeit verwendet die Interpunktionssetzung als Indikator für Haupt- und Nebensätze. Die Satzzeichen Punkt, Fragezeichen und Ausrufezeichen – in Kombination mit dem jeweiligen Satzbeginn – signalisieren Hauptsätze, während die Nebensätze anhand der

Satzzeichen Komma, Klammer, Gedankenstrich und Strichpunkt bestimmt werden. Formel zur Berechnung:

$$Subordinationsindex\ (SOI) = \frac{\sum Hauptsätze}{\sum Nebensätze}$$

4. Formel: Die Berechnung des Subordinationsindexes (vgl. Schubert 2008: 178)

3.5 Dogmatismusquotient: Dogmatische A- und B-Ausdrücke als Indikatoren für die Komplexitätsreduzierung und Komplexitätstoleranz

Der Dogmatismusbegriff stammt ursprünglich von Rokeach, der ihn im Rahmen seiner Autoritarismusforschung in den 1960er-Jahren eingeführt hat (vgl. Schubert 2008: 112). Im Mittelpunkt der Aufmerksamkeit stehen nicht die Inhalte von Überzeugungssystemen, sondern die sprachliche Struktur, die sie bildet. Die zwei Pole des Indikators skalieren die jeweilige Offenheit bzw. Geschlossenheit des analysierten Gedankensystems. Laut Rokeach unterscheiden sich Menschen in zwei Bedürfnissen. „Das Bedürfnis nach einem kognitiven Bezugsrahmen zur Orientierung in der Welt" interpretiert er als Offenheit (Schubert 2008: 113), während „das Bedürfnis, bedrohliche Aspekte der Umwelt abzuwehren", von ihm als Geschlossenheit interpretiert wird (Schubert 2008: 113).

Rokeach geht von einer Opposition zwischen ‚Beliefs' und ‚Disbeliefs' aus. In die erste Kategorie fallen Überzeugungen, die als wahr akzeptiert werden, während zur zweiten Kategorie jene Instanzen gehören, die als falsch bzw. ungültig abgelehnt werden (vgl. Schubert 2008: 113). „Jede Information wird dahingehend überprüft, ob sie mit den zentralen Grundüberzeugungen übereinstimmt. Informationen, die nicht in das Überzeugungssystem integrierbar sind, werden abgewehrt oder verzerrt" (Schubert 2008: 113). Ertel interpretiert die bevorzugten Gruppen jener Verwendungen als Sprechhandlungstypen, für die er ein Verfahren zur Dogmatismustextauswertung entwickelt. Er hypothetisiert unterschiedliche Gruppen von Wörtern, die auf der Sprachoberfläche den jeweiligen Grad der Offenheit bzw. Geschlossenheit des Gedankensystems des Sprachbenutzers widerspiegeln (vgl. Schubert 2008: 113).

Roth klassifiziert u.a. den Gebrauch von „Generalisierungen", die „Formulierung von Ausschlußbeziehungen" sowie Gewissheits- bzw. Notwendigkeitsausdrücke als „geschlossenheitsindizierend" (vgl. Schubert 2008: 114). Im DOTA-Kodierlexikon listet er insgesamt 433 lexikalische Einheiten auf, die sechs Subkategorien zugeordnet sind. Die Auftretenshäufigkeit der „A-Ausdrücke" wird als „Tendenz zur komplexitätsreduzierenden Informationsverarbeitung" interpretiert, während die Anzahl der „B-Ausdrücke" eine „Tendenz, Komplexität, Ungewissheit und Unvollkommenheit zu tolerieren", zum Ausdruck

bringt (Schubert 2008: 114). Die häufigsten Kategorien der A- und B-Ausdrücke im DOTA-Kodierlexikon sind die folgenden lexikalischen Einheiten:

A-Ausdrücke	B-Ausdrücke
Kategorie 1 (Häufigkeit, Dauer und Verbreitung): beständig, immer, jederzeit, jedesmal, nie, niemals, ständig, stets, allemal, endgültig.	Kategorie 1 (Häufigkeit, Dauer und Verbreitung): ab und zu, im Allgemeinen, gelegentlich, gewöhnlich, häufig, hin und wieder, mehrfach, meistens, mitunter, normalerweise.
Kategorie 2 (Anzahl und Menge): alle, ausnahmslos, ohne Einschränkung, einzig, ganz, nicht im geringsten, gesamt, jede, jedermann, jegliche.	Kategorie 2 (Anzahl und Menge): eine ~ Anzahl, ein bisschen, einzelne, etwas, gewisse, größtenteils, mehrere, eine Menge, ein paar, teilweise
Kategorie 3 (Grad und Maß): absolut, gänzlich, ganz und gar, grundlegend, grundsätzlich, von Grund auf, in vollem Maße, prinzipiell, restlos, total.	Kategorie 3 (Grad und Maß): besonders, einigermaßen, im ~ Grade, höchst, kaum, mehr oder minder, relativ, sehr vorwiegend.
Kategorie 4 (Gewissheit): ausgeschlossen, eindeutig, einwandfrei, fraglos, gewiss, nicht im mindestens, natürlich, notwendig, sicher, mit Sicherheit.	Kategorie 4 (Gewissheit): allenfalls, dem Anschein nach, augenscheinlich, denkbar, fraglich, immerhin, möglich, mutmaßlich, offenbar.
Kategorie 5 (Ausschluss, Einbeziehung und Geltungsbereich): allein, alles andere als, ausschließlich, einzig, allein, entweder oder, lediglich, nichts als, nichts weiter, nur, weder noch.	Kategorie 5 (Ausschluss, Einbeziehung und Geltungsbereich): unter anderem, andererseits, auch, außerdem, darüber hinaus, ebenfalls, zum einen, einerseits, einschließlich, ferner.
Kategorie 6 (Notwendigkeit und Möglichkeit): müssen, nicht dürfen, nicht können, sich nicht lassen, nicht -bar sein, nicht in der Lage sein, nicht imstande sein.	Kategorie 6 (Notwendigkeit und Möglichkeit): dürfen, können, sich lassen, -bar sein, in der Lage sein, vermögen, nicht brauchen zu, nicht müssen.

Tabelle 1: Die dogmatischen A- und B-Ausdrücke nach den 6 Kategorien (Schubert 2008: 114)

Bei Schubert erscheint der Dogmatismusquotient als Indikator für die Geschlossenheit des Denksystems (vgl. Schubert 2008: 178). Die Formeln sind die folgenden:

$$Dogmatismusquotient\ (DQ) = \frac{\sum D(A)}{\sum D(A) + \sum D(B)}$$

4. Formel: Die Berechnung des Dogmatismusquotienten (vgl. Schubert 2008: 178)

Da der Dogmatismusquotient eine Offenheit/Geschlossenheit-Skala hypothetisiert, sind dogmatische A-Ausdrücke ein Indikator für Komplexitätsreduzierung, während B-Ausdrücke den Grad der jeweiligen Komplexitätstoleranz anzeigen:

$$D(A) = \frac{\sum D(A)}{Textlänge}$$

6. Formel: Die Berechnung der relativen Anzahl der dogmatischen A-Ausdrücke (vgl. Schubert 2008: 178)

$$D(B) = \frac{\sum D(B)}{Textlänge}$$

7. Formel: Die Berechnung der relativen Anzahl der dogmatischen B-Ausdrücke (vgl. Schubert 2008: 178)

3.6 Referentielle und Operative Prägnanz als Indikatoren für die Tendenz zur Generalisierung und für die Kohärenz des Denksystems

Die referentielle Prägnanz lässt sich als Indikator für die Tendenz zur Generalisierung betrachten, während die operative Prägnanz als Indikator für die Kohärenz des Denksystems gedeutet wird (vgl. Schubert 2008: 178). Ertel bezieht im Rahmen des Dogmatismuskonzepts gestalttheoretische Überlegungen ein (vgl. Schubert 2008: 115). Nach Metzger können wir unter dem Prägnanzprinzip eine Fähigkeit von Organismen verstehen, „die bestmögliche Wahrnehmungsorganisation herzustellen“ (Schubert 2008: 115). Rausch unterscheidet zwischen sieben bipolaren Prägnanzdimensionen. Im Gegensatz zu Gesetzmäßigkeit, Eigenständigkeit, Integrität, Einfachheit, Komplexität, Ausdrucksstärke und Bedeutungsfülle stehen die Dimensionen von Zufälligkeit, Abgeleitetheit, Privativität, Kompliziertheit, Tenuität, Ausdrucksschwäche und Bedeutungsschwäche (vgl. Schubert 2008: 115).

Rausch interpretiert die Dimension der Einfachheit im gestalttheoretischen Rahmenkonzept so, dass die Einfachheit als solche nicht anders ist als die Bevorzugung eines geordneten Gebildes (aus der Menge aller Gebilde), dessen Gesetzmäßigkeit – oder Geordnetheit – leicht zu erkennen ist (Schubert 2008: 115). Im Gegensatz zur Geordnetheit, in der die Gesetzmäßigkeiten erscheinen sollen, steht der Zustand der Zufälligkeit, d.h. der der Entropie. Die Dimension der Komplexität deutet er als „gedrängte, klare und scharfe Formulierung“ (zit. n. Schubert 2008: 115). Die sieben Merkmalsdimensionen nehmen am gemeinsamen Moment der „Mitbestimmtheit durch das Subjekt“ und der „Konstruktivität“ teil (zit. n. Schubert 2008: 116).

Witte definiert die Gestalt als „gegliedertes Ganzes“ bzw. als „wohlstrukturierte Gliederung“, die sich in psychischen Bezugssystemen manifestiert. Laut Metzger sind die Denkprozesse mit den Prägnanztendenzen vergleichbar, wie es auch in der Prototypenlehre bzw. in der Semantiktheorie erscheint (zit. n. Schubert 2008: 116). Elemente der Wahrnehmung, die mit dem Bezugssystem der jeweiligen Gestaltidentifizierung nicht kohärent sind, werden aus dem Mittelpunkt der Aufmerksamkeit gebannt (vgl. Schubert 2008: 117). Ertel meint, der „DQ eines Textes – so lautet jetzt die einheitsstiftende Interpretation – repräsentiert das Prägnanzniveau der kognitiven Tätigkeit, das der Produktion der jeweils untersuchten Textmenge zugrunde liegt“ (Schubert 2008: 118f.).

Schwibbe bzw. Ertel haben erkannt, dass der Dogmatismusquotient faktorenanalytisch in zwei eigenständige Faktoren getrennt werden kann (vgl. Schubert 2008: 119). Die Kategorien 1, 2, 3 und 5 bilden den D1235-Faktor, während der D46-Faktor aus den Kategorien 4 und 6 gebildet wird. Während der erste den Geltungsbereich der Aussagen bestimmt, definiert der zweite die

Kohärenz gedanklicher Abläufe (vgl. Schubert 2008: 119). Der D1235-Faktor misst die Verwendung generalisierender Lexeme, die eine Komplexitätsreduzierung bzw. Redundanzbildung zum Ausdruck bringen, während die Kategorien des D46-Faktors die Parameter angeben, nach denen laut Schwibbe die sprachliche Fertigkeit modelliert werden kann, „Einzelinformation in logische Relationen setzen zu können" (Schubert 2008: 119). Roth meint, dass der „Variation von Teilindikatoren […] unterscheidbare Aspekte der kognitiven Dynamik zugrundeliegen" (zit. n. Schubert 2008: 119). Formel zur Berechnung:

$$Referentielle\ Prägnanz\ (D1235) = \frac{(D[A1]+D\ [A2]+D\ [A3]+D\ [A5])}{(D\ [A1]+D\ [A2]+D\ [A3]+D\ [A5]+D\ [B1]+D\ [B2]+D\ [B3]+D\ [B5]}$$

8. Formel: Die Berechnung der referentiellen Prägnanz (vgl. Schubert 2008: 178)

$$Operative\ Prägnanz\ (D46) = \frac{(D[A4]+D\ [A6])}{(D\ [A4]+D\ [A6]+D\ [B4]+D\ [B6]}$$

9. Formel: Die Berechnung der operativen Prägnanz (vgl. Schubert 2008: 178)

3.7 Konjunktive als Indikatoren für die Vorstellungskraft

Der Gebrauch des Konjunktivs kann als Indikator für Vorstellungskraft verstanden werden (vgl. Schubert 2008: 178). Laut Roth bezeichnet die Konjunktivverwendung „das Verhältnis der thematisierten Sachverhalte zur außersprachlichen Realität" (zit. n. Schubert 2008: 142). Schwibbe sieht in der Verwendung von Konjunktiven das Imaginieren von „hypothetischen Geschehnissen und fiktionalen Relationen" (zit. n. Schubert 2008: 142). Sie betrachtet die jeweilige Konjunktivverwendung als Indikator für die zum Ausdruck gebrachte Vorstellungskraft. Erben interpretiert die Verwendung von Konjunktiven als „das Erbauen und Erleben einer fiktiven Welt auf der Experimentierbühne des Konjunktivs" – als ein Spiel mit Möglichkeiten (zit. n. Schubert 2008: 142). Laut Schöne ermöglicht das konjunktivische Denken das Hypothetisieren als solches bzw. die Überprüfung von Hypothesen (vgl. Schubert 2008: 143). Aus forschungspraktischen Gründen wurden in der vorliegenden Arbeit nur die Konjunktiv-II-Formen der Modal- und Hilfsverben berücksichtigt. Formel zur Berechnung der relativen Konjunktivzahl:

$$Konjunktive = \frac{\sum Konjunktive}{Textlänge}$$

10. Formel: Die Berechnung der relativen Anzahl der Konjunktive (vgl. Schubert 2008: 178)

3.8 Negationen als Indikatoren für die Vermeidung von Ambiguität

Negationen können als Indikator für die Vermeidung von Ambiguität interpretiert werden (vgl. Schubert 2008: 178). Roth interpretiert den vermehrten Gebrauch von Negationen als eine Tendenz, inkonsistente Informationen abzuwehren (Schubert 2008: 152). In der vorliegenden Arbeit wurden die folgenden Negationen in die Berechnung einbezogen: ‚nicht', ‚kein', ‚nirgendwo' und ‚nie'. Formel zur Berechnung der relativen Anzahl von Negationen:

$$Negationen = \frac{\sum Negationen}{Textlänge}$$

11. Formel: Die Berechnung der relativen Anzahl der Negationen (vgl. Schubert 2008: 178)

3.9 Begründungen und Entgegensetzungen als Indikatoren für die Verknüpfungsfähigkeit und für die Abgrenzung

Begründungen können die Verknüpfungsfähigkeit signalisieren (vgl. Schubert 2008: 178). Laut Räder dienen Konjunktionen „der Verknüpfung von Aussagen" (Schubert 2008: 151). Räder sieht Konjunktionen als Indikatoren syntaktischer Komplexität (Schubert 2008: 151). Schwibbe systematisiert die Funktionen von Konjunktionen mit den folgenden Worten:

> Allgemein können wir von diesen Konstruktionselementen sagen, daß sie generalisierte Informations- und Bedeutungsträger sind, die dazu dienen, spezifische Bedeutungselemente zu einem sinnvollen sprachlichen Kontext zu organisieren. Im Zusammenhang damit können wir sie auch als generalisierte ‚Funktoren' interpretieren, die auf ökonomische Weise Unsicherheiten in einer Mitteilung abbauen, daß heißt Information herstellen. (Zit. n. Schubert 2008: 151)

Laut Roth stellen Konjunktionen einen kausalen Zusammenhang zwischen Sachverhalten her (Schubert 2008: 151). Die Häufigkeit der verwendeten Begründungen zeigt, inwieweit der Sprachbenutzer bemüht ist, getroffene Einzelaussagen „in einen inhaltlich-logischen Zusammenhang einzubinden" (Schubert 2008: 335). In der vorliegenden Arbeit wurden in die Analyse die Wörter ‚weil', ‚da', ‚damit', ‚zumal' und ‚also' als Begründungen einbezogen. Während die Konjunktionen ‚weil', ‚da' und ‚damit' eindeutig als Begründungswörter gelten, können ‚zumal' und ‚also' auch eine begründende Absicht implizieren. Formel zur Berechnung:

$$Begründungen = \frac{\sum Begründungen}{Textlänge}$$

12. Formel: Die Berechnung der relativen Anzahl der Begründungen (vgl. Schubert 2008: 178)

Entgegensetzungen können als Indikator für Abgrenzung fungieren (vgl. Schubert 2008: 178). In der vorliegenden Arbeit umfasst die Kategorie der Entgegensetzungen die Berechnung der Wörter ‚aber', ‚sondern', ‚stattdessen', ‚trotzdem', ‚hingegen' und ‚jedoch', d.h. Wörter, die explizit oder implizit eine adversative konjunktivische Funktion haben können. Nach Roth

beschreiben adversative Konjunktionen „Unstimmigkeiten und Gegensätze“ (Schubert 2008: 151). Laut Schwibbe besteht die Funktion von Entgegensetzungen darin, „das eigene Überzeugungssystem gegen andere Systeme und gegen labilisierende Informationen abzugrenzen“ (zit. n. Schubert 2008: 151). Formel zur Berechnung der relativen Anzahl der Entgegensetzungen:

$$Entgegensetzungen = \frac{\sum Entgegensetzungen}{Textlänge}$$

13. Formel: Die Berechnung der relativen Anzahl der Entgegensetzungen (vgl. Schubert 2008: 178).

3.10 Modalverben als Indikatoren für das Einbeziehen externer Randbestimmungen

Der Gebrauch der Modalverben lässt sich als Indikator für das Einbeziehen externer Randbedingungen interpretieren (vgl. Schubert 2008: 178). Modalverben modifizieren den Inhalt bzw. die Geltungsmodalität anderer Verben: Notwendigkeit, Wille, Möglichkeit, Wunsch, Fähigkeit oder Ungewissheit (vgl. Schubert 2008: 144). In der vorliegenden Arbeit wurden aus forschungspraktischen Gründen die sechs klassischen Modalverben nach Kipper (zit. n. Schubert 2008: 144) in ihren verschiedenen Erscheinungsformen berechnet. „Modalverben haben eine kommunikative Funktion, indem der Sprecher durch sie sein Wissen, seine Handlungen und Handlungsabsichten in Relation zu denen anderer setzt“ (Schubert 2008: 144). Formel zur Berechnung der relativen Anzahl der Modalverben:

$$Modalverben = \frac{\sum Modalverben}{Textlänge}$$

14. Formel: Die Berechnung der relativen Anzahl der Modalverben (vgl. Schubert 2008: 178)

3.11 Die relative Textlänge als Indikator für die Informationsmenge

Der Indikator der relativen Textlänge ermöglicht den Vergleich der Textlänge des gegebenen Textes mit anderen. Die Division setzt die Textlänge ins Verhältnis zur Zahl der analysierten Texteinheiten (vgl. Schubert 2008: 178). Es besteht die Wahrscheinlichkeit, dass ein hoher Wert der relativen Textlänge als Merkmal der parallel-holistischen Informationsverarbeitung betrachtet werden kann (vgl. Kuhl/Quirin/Koole 2020: 10), da die jeweilige Textlänge die „Verarbeitungsleistung von großen Informationsmengen“ qualifiziert (Schubert 2008: 176). Formel zur Berechnung der relativen Textlänge:

$$relative\ Textlänge = \frac{Textlänge}{\sum Zahl\ der\ Texteinheiten}$$

15. Formel: Die Berechnung der relativen Textlänge

3.11 Hilfsverben als vermutete Indikatoren für die Schattierung von Modalitäten

Die Bedeutung der relativen Anzahl von Hilfsverben im persönlichkeitsspezifischen Sprachgebrauch ist noch nicht ausreichend erforscht. Man kann jedoch davon ausgehen, dass es in der Verwendung von Hilfsverben zwischen Informationsverarbeitungsgruppen einen grundlegenden Unterschied gibt (vgl. Schubert 2008: 177). Formel zur Berechnung:

$$Hilfsverben = \frac{\sum Hilfsverben}{Textlänge}$$

16. Formel: Die Berechnung der relativen Anzahl der Hilfsverben (vgl. Schubert 2008: 178)

4. Darstellung des Forschungsablaufs und der Ergebnisse

Im Mittelpunkt der Untersuchung steht die Beschreibung der Eigenschaften statistischer Sprachgebrauchsmuster, anhand derer kognitive Merkmale der zugrunde liegenden Informationsverarbeitungssysteme erschlossen werden können. Allerdings wird die Situativität bewusst ausgeklammert. Die Arbeit geht davon aus, dass sich der situative Einfluss durch die Erhöhung der analysierten Texteinheiten ausschalten bzw. reduzieren lässt. Da Messwerte seltenerer Situationen die großen Durchschnittswerte nicht maßgeblich beeinflussen, bleibt die Gesamtbewertung stabil. Zudem kann die Situation als eine kognitive Konstruktion betrachtet werden, die vielmehr eine subjektive, individuumsspezifische Größe darstellt.

Auf Grund der statistischen Ergebnisse wurden die Funktionsprofile der vorherrschenden Informationsverarbeitungs- bzw. Persönlichkeitsstile ermittelt (vgl. Schubert 2008: 178). Dafür wurden sechs Indikatoren ausgewählt, die Schubert u.a. als die am besten abgesicherten Indikatoren einstuft. Zur Berechnung des OE/IV-Anteils dienten der Variationsindex, der Subordinationsindex sowie die relative Textlänge. Für die Berechnung des GF/AD-Anteils wurden der Abstraktheitsindex sowie die Häufigkeit dogmatischer A- und B-Ausdrücke herangezogen. Die Ermittlung der jeweiligen Stile basiert auf einem Interpretationsschritt, der sich an den Empfehlungen und Annahmen Schuberts orientiert. Um den Geltungsbereich der berechneten Stile zu bestimmen, sind jedoch weitere Forschungen erforderlich.

4.1 Darstellung der Auswahl der Probanden

Im Rahmen des Korpusbildungsprozesses wurden insgesamt 10.000 Blogbeiträge von der Webseite GuteFrage.de analysiert. Diese stammen von 55 zufällig ausgewählten Sprachbenutzern. Fünf von ihnen (Probanden 1–5) lieferten jeweils 1.000 Blogbeiträge. Ziel war es, Kohärenzmuster der Indikatorwerte anhand der primären Verteilung der Beiträge zu ermitteln. Zusätzlich wurden jeweils 100 Blogbeiträge von 50 weiteren Sprachbenutzern (Probanden 6–55) ausgewertet, um den Bewegungsraum der Indikatorwerte zu erfassen. Bei

der Auswahl der Probanden spielte ihre Aktivität auf Gutefrage.de eine wesentliche Rolle. Berücksichtigt wurden Nutzer, die in den letzten Jahren eine hohe Aktivität zeigten, erkennbar u.a. an einer öffentlich zugänglichen Topliste der aktivsten Benutzer. Diese Nutzer wurden in die sprachprofilierende Analyse einbezogen. Die Benutzer mit jeweils 100 Blogbeiträgen erhielten eine Nummer zwischen 1 und 50, während die Probanden mit jeweils 1.000 Texteinheiten mit einer Nummer zwischen 1 und 5 versehen wurden.

4.2 Darstellung der Korpusbildung

Die Blogbeiträge wurden zufällig ausgewählt. Berücksichtigt wurden jeweils die letzten 1.000 bzw. 100 Beiträge der einzelnen Sprachbenutzer. Die 10.000 Blogbeiträge wurden manuell gespeichert und anschließend in Python mit Zeilennummern versehen. Dies ermöglichte eine präzise analytische Verteilung der Texteinheiten. Die 55 Korpora (6.000 A4-Seiten, Schriftgröße 12) wurden im .docx-Format gespeichert und mit MS Word verwaltet. Um Ausnahmen innerhalb der einzelnen Suchkategorien der sprachstatistischen Indikatoren zu identifizieren, wurden erste Untersuchungen als Stichproben durchgeführt. Diese erfolgten manuell mithilfe des Suchmotors von MS Word, wobei Regex-Wortformate auf das Gesamtkorpus angewendet wurden. Dadurch konnten potenzielle Fehler bei der Suffixfrequenzsuche eliminiert werden. Die sprachstatistischen Indikatoren basieren größtenteils auf Schuberts Beschreibung. Einige Anpassungen wurden vorgenommen, um die Datenvisualisierung zu optimieren. Die Analysewerte des Python-Codes wurden mehrfach mit den manuellen Stichprobenergebnissen verglichen, um mögliche Fehler in der Automatisierung zu erkennen und zu korrigieren.

Die Rohdaten der softwaregestützten Analyse wurden in MS Excel gespeichert und ausgewertet. Ein Vorteil der zufällig ausgewählten Blogbeiträge liegt in ihrer Spontaneität. Es handelt sich um Texte, die ohne Einfluss einer Beobachtungssituation verfasst wurden. Der Leitgedanke der Forschung basiert auf der Annahme, dass die Verfasser ihren natürlichen ‚Bedürfnissen' folgen. Sie verfassen spontane Freizeitblogbeiträge. Dadurch spiegeln die Ergebnisse keine künstliche Thematik oder ein künstliches Sprachregister wider. Stattdessen werden die natürlichen sprachlichen Persönlichkeitsmerkmale der Verfasser sichtbar. Ein weiterer wichtiger Aspekt der Forschung ist die Situation der Textverfassung. Der Verfasser sitzt allein in einem Zimmer vor dem Computer und konstruiert die Schreibsituation für sich selbst. Daher unterliegen die Blogbeiträge keinem prägenden situativen Einfluss. Die Situation ist vielmehr eine kognitive Konstruktion. In jedem Fall folgt die Verfassung der Texte dem eigenen Interesse oder Bedürfnis. Sie dient der Externalisierung gewisser kognitiver Muster.

Von den sprachstatistischen Indikatoren nach Schubert wurden insgesamt 16 ausgewählt. Ziel war es, die kognitiven Informationsverarbeitungsmodi analytisch einzugrenzen. Emotionsanalytische Kategorien blieben unberücksichtigt. Die zugrunde liegenden Emotionswörterbücher stammen aus der zweiten Hälfte des 20. Jahrhunderts und sind in ihrer heutigen Gültigkeit umstritten. Um die Situativität ausreichend auszuschalten, wurde eine hohe Anzahl von Texteinheiten einbezogen. Die Untersuchung geht davon aus, dass sich häufige und seltene Situationen aus dem kognitiven Funktionieren des Einzelnen konstruieren.
Dementsprechend kann die Situativität in der Analyse weitgehend ausgeschlossen werden. Eine Situationsforschung ist für diese Untersuchung nicht erforderlich. Die große Anzahl der Texteinheiten sorgt dafür, dass die Analysewerte häufiger auftretender Situationen die Verteilungswerte seltenerer Situationen korrigieren. Um die Situativität ausreichend auszuschalten, erwies sich eine primitive Verteilung als ausreichend. Diese umfasste 1×1000, 2×500 und 10×100 Texteinheiten bei den fünf Sprachbenutzern mit jeweils 1000 Analyseeinheiten. Zusätzlich wurde eine Verteilung von 1×100 und 2×50 Texteinheiten bei den 50 Verfassern mit jeweils 100 Analyseeinheiten angewendet. Die Datenvisualisierung wurde in MS Excel durchgeführt. Um eine klare Darstellung zu gewährleisten, wurden die Rohwerte anhand der Peak-/Bottom-Werte der 50 Sprachbenutzer auf einer Skala von 0 bis 1 normiert. Dadurch bewegen sich alle Indikatorwerte auf derselben Skala zwischen vergleichbaren Gegenpolen.

4.3 Darstellung der Roh- und Visualisierungsdaten

Die darzustellenden Sprachprofilierungen vermitteln ein kohärentes Bild vom Sprachgebrauch des betreffenden Sprachbenutzers. Die Sprachgebrauchsmuster erwiesen sich als äußerst kohärent. In Zukunft könnte die Einbeziehung einer tiefgehenden Situations- bzw. Variationsanalyse lohnenswert sein, um herauszufinden, welche Muster sich jenseits der drastischen Vereinfachung einfacher Verteilungen auszeichnen und ob sowie wie sich die Sprachgebrauchsmuster bestimmten Situationsmustern zuordnen lassen. Mit den unterschiedlichen innovativen Fragestellungen rund um die Sprachgebrauchsmuster zeichnet sich ein robustes Thema im Sinne daktyloskopischer Spurelemente in sprachlichen Strukturen ab, von denen keine Ebene der menschlichen Psyche und Sprachvariation unberührt bleibt. Die heutigen Entwicklungstendenzen der Computerlinguistik geben ein hoffnungsvolles Bild davon, ob sich jene äußerst feinen (quantischen) Spurelemente in den nächsten Jahrzehnten entschlüsseln lassen (vgl. Müller 2013: 183).

4.3.1 Bewegungsraum der Indikatorwerte bei den Probanden 6–55

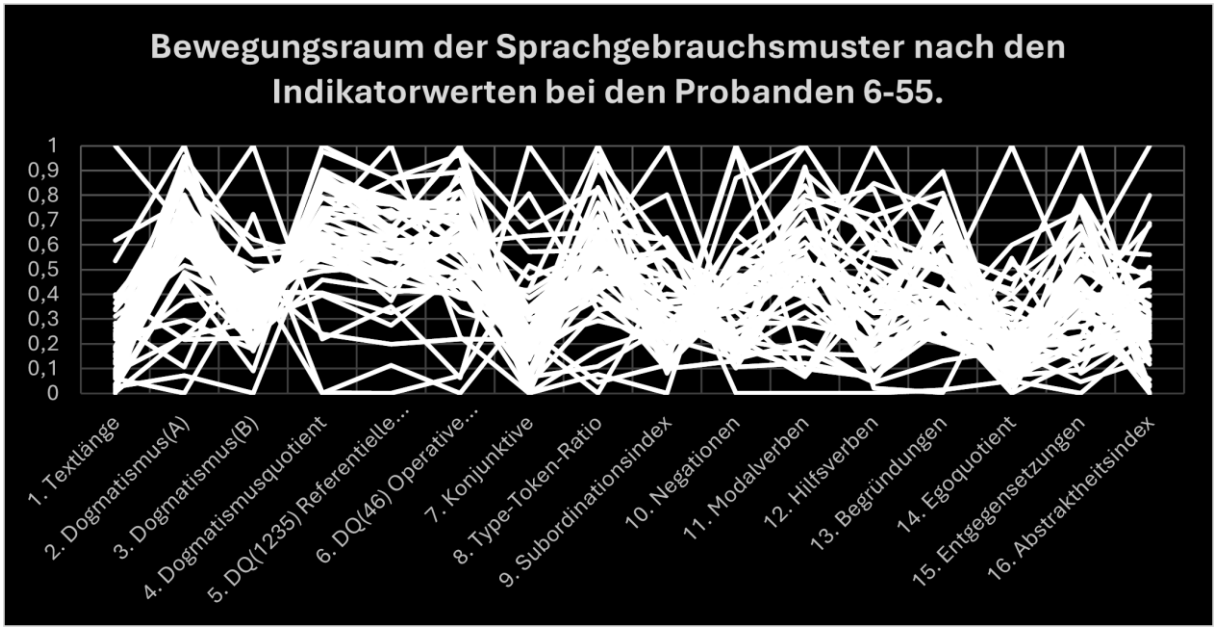


Abb. 5: Bewegungsraum der Sprachgebrauchsmuster anhand der Visualisierungswerte der 16 sprachstatistischen Indikatoren bei den Probanden 6–55

Abb. 5 zeigt den Bewegungsraum der Visualisierungswerte der Sprachgebrauchsmuster. Die Darstellung basiert auf 16 sprachstatistischen Indikatoren. Die Werte liegen zwischen 0 und 1, sodass ihre Normal- bzw. Durchschnittskurven erkennbar sind. Die Normalkurven konzentrieren sich größtenteils auf die Mittelebene, wie es statistisch zu erwarten ist. Seltener auftretende über- und unterdurchschnittliche Werte befinden sich an den Rändern. Der in Abb. 5 sichtbare Bewegungsraum der Indikatorwerte bildet den jeweiligen statistischen Kontext der Sprachgebrauchsmuster. Diese werden in den folgenden Unterkapiteln näher erläutert. Allerdings basiert die Berechnung auf insgesamt 50 Probanden und 5000 Blogbeiträgen. Damit erfüllt der dargestellte Bewegungsraum nur eine minimale Voraussetzung für statistische Repräsentativität. Das Ergebnis deutet darauf hin, dass ein Data-Science-Design mit mehreren tausend Probanden in Zukunft aussagekräftigere Erkenntnisse über die Verteilung der Indikatorwerte liefern könnte.

4.3.2 Sprachgebrauchsmuster nach den Indikatorwerten beim Probanden 1

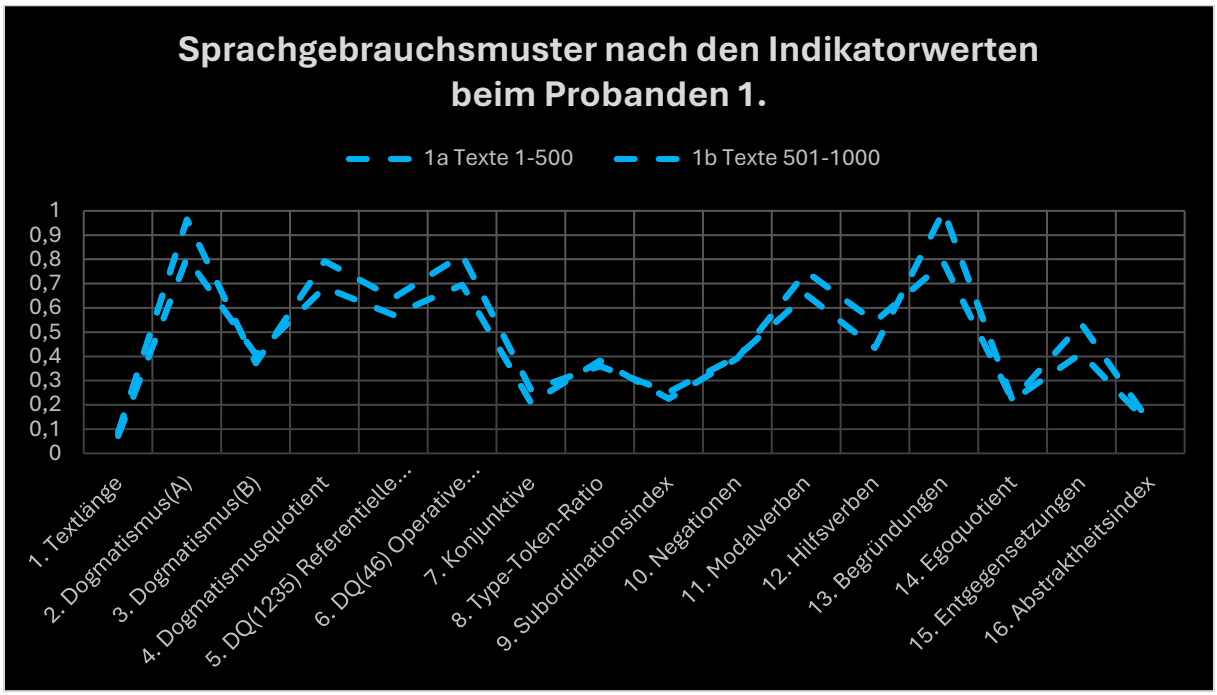


Abb. 6: Signifikante Übereinstimmungen der Sprachgebrauchsmuster beim Probanden 1

Bei Abb. 6 und 7 werden die Ergebnisse von Proband 1 dargestellt. Abb. 6 zeigt die Indikatorwerte von Proband 1, dessen Texteinheiten durch eine einfache Verteilung auf die erste und zweite Hälfte der analysierten Blogbeiträge aufgeteilt wurden. Die Texteinheiten von 1–500 bzw. 501–1000 werden durch die Vergleichsgruppen 1a und 1b gekennzeichnet. Die Abb. verdeutlicht, dass die Sprachgebrauchsmuster relativ konstant sind und keine statistisch signifikanten Abweichungen aufweisen – wie es bei der Darstellung der Z-Test-Ergebnisse detaillierter erläutert wird. Daraus lässt sich schließen, dass die Indikatorwerte überzufällige Sprachgebrauchsmuster sind, die ein bestimmtes neuronales Funktionieren signalisieren und sich statistisch auf der Sprachoberfläche manifestieren.

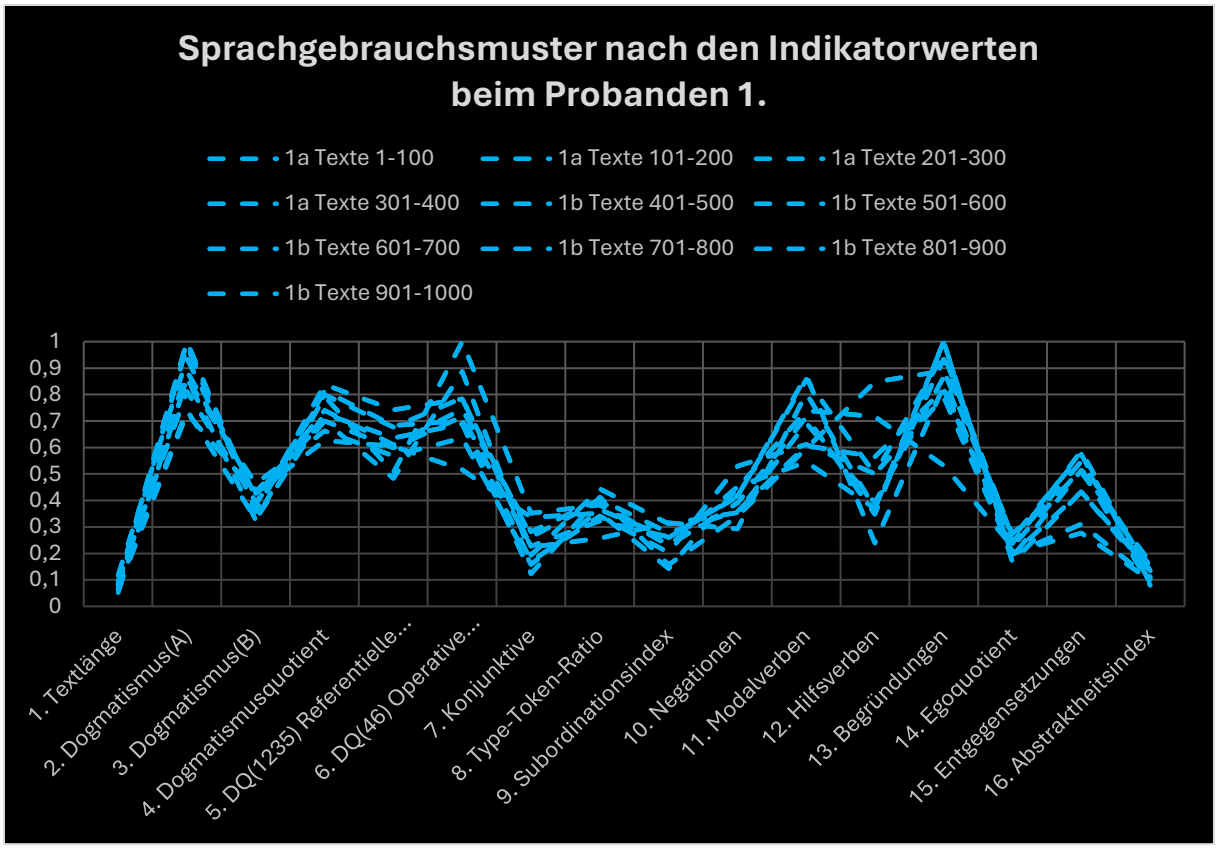


Abb. 7: Tendenzen der Sprachgebrauchsmuster bei dem Probanden 1

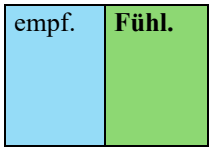
Abb. 7 zeigt, dass die Tendenzen, die durch die einfache Verteilung der Texteinheiten in die Intervalle 1–100, 101–200, 201–300, 301–400, 401–500, 501–600, 601–700, 701–800, 801–900 und 901–1000 entstehen, ebenfalls relativ konstant bleiben. Dies verstärkt die Annahme, dass Sprachgebrauchsmuster überzufällig auftreten. Allerdings ist festzustellen, dass die Korpora potenzielle Störfaktoren enthalten können. Zitate, Links, Rückverweise auf bereits Erwähntes sowie Primingeffekte auf der Sprachoberfläche könnten die Ergebnisse beeinflussen. Obwohl die Einbeziehung einer relativ großen Anzahl von Blogbeiträgen offenbar ausreicht, um situative Einflüsse auszuschließen, zeigt sich dennoch, dass die Aufteilung der Texteinheiten im Verhältnis 500:500 eine deutlich höhere Genauigkeit aufweist als die im Verhältnis 100:100.

Basierend auf den bereits erwähnten Annahmen Schuberts (Abb. 3–4) sowie den ausgewählten Indikatorergebnissen lassen sich für Proband 1 die folgenden dominierenden Funktionen der menschlichen Informationsverarbeitung und die möglichen Persönlichkeitsstile hypothetisieren. Laut Schubert korrelieren der Subordinationsindex, der Variationsindex und die Textlänge positiv mit dem ganzheitlich Informationsverarbeitungsstil und negativ mit dem analytisch-sequenziellen Stil. Im Gegensatz dazu korrelieren der Abstraktheitsindex und die

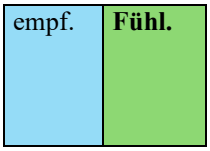
dogmatischen B-Ausdrücke negativ mit dem ganzheitlichen Stil und positiv mit dem analytisch-sequenziellen Stil. Die dogmatischen B-Ausdrücke zeigen dabei entgegengesetzte Tendenzen. Die ausgewählten Indikatoren wurden als Interpretationsschritt den zugehörigen kognitiven Funktionen zugeordnet, wie in Abb. 8 ersichtlich ist. Die Inhalte der Indikatoren – darunter Komplexitätsreduzierung, Komplexitätstoleranz, Verdichtungen im Denksystem, Informationsmenge, Wortschatz und der Umgang mit komplexen Satzeinheiten (Abb. 3) – können als weitgehend unabhängige Funktionen des jeweiligen Informationsverarbeitungssystems betrachtet werden. Obwohl die ersten drei Indikatoren für die AD/GF-Achse verwendet wurden und die letzten drei für die IV/OE-Achse, scheinen sie dennoch relativ unabhängig voneinander zu sein.

Es kann vermutet werden, dass die jeweiligen Teilsysteme (AD, GF, IV und OE) als Funktionen der Komplexitätsreduktion und anderer kognitiver Prozesse wirken. Dennoch bleiben sie relativ unabhängig voneinander. Trotz dieser angenommenen Unabhängigkeit sind sie jedoch stark personenspezifisch. Anhand von Abb. 4 wurden die Verhältnisse der Teilsysteme den jeweiligen Persönlichkeitsstilen zugeordnet. Schubert grenzt nicht eindeutig ab, welche Persönlichkeitsstile daraus abgeleitet werden können. Sie stellt lediglich die Konstellation der dominanten Funktionen dar, was darauf hindeutet, dass die jeweiligen Stile eine strukturelle Pluralität aufweisen könnten. Weitere Untersuchungen sind erforderlich, um die explorierten Stile genauer zu bestimmen. Die Pluralität der vermuteten Persönlichkeitsstile verweist auf ein bekanntes Problem der Semantik: Die Ergebnisse erweisen sich im Rahmen der verwendeten Theorie als polysem. Bei Proband 1 dominieren – basierend auf den ausgewählten Indikatoren – die Funktionen des Empfindens (OE) und Fühlens (GF). Daraus lässt sich ableiten, dass entweder eine Variation des selbstbestimmten oder des sorgfältigen sprachlichen Persönlichkeitsstils vorliegt.

Proband 1	Empfindendes	Fühlen			
Indikator	Assoziierte Funktion	%	Indikator	Assoziierte Funktion	%
Subordinationsindex	IV (+) OE (-)	0.24	Abstraktheitsindex	AD (+) GF (-)	0.12
Textlänge	IV (+) OE (-)	0.07	Dogmatisches A	AD (-) GF (+)	0.89
Variationsindex	IV (+) OE (-)	0.37	Dogmatisches B	AD (+) GF (-)	0.38
Dominante Funktion	OE (88%) IV (22%)	OE	Dominante Funktion	GF (80%) AD (20%)	GF



1. Selbstbestimmt b)



5. sorgfältig b)

Abb. 8: Die Ergebnisse der ausgewählten Indikatoren und die assoziierten Funktionen bzw. vermuteten Persönlichkeitsstile beim Probanden 1

4.3.3 Sprachgebrauchsmuster nach den Indikatorwerten beim Probanden 2

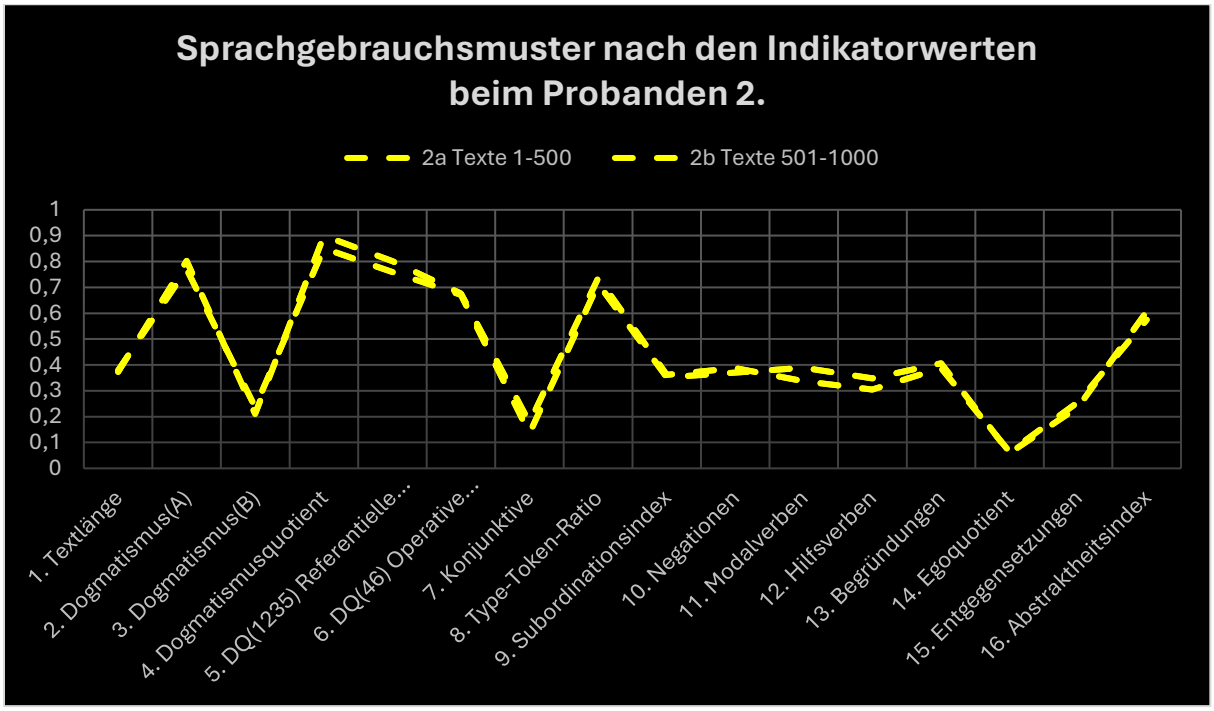


Abb. 9: Signifikante Übereinstimmungen der Sprachgebrauchsmuster beim Probanden 2

In Abb. 9 sind die Ergebnisse bei Probanden 2 nach der zuvor erläuterten Systematik dargestellt. Die Abb. verdeutlicht, dass die Vergleichsgruppen 2a und 2b, bestehend aus den Texteinheiten 1–500 bzw. 501–1000, keine signifikanten Abweichungen aufweisen. Auch Abb. 10 stützt die Annahme der Überzufälligkeit der Sprachgebrauchsmuster, da die Tendenzen der

Indikatorwerte weitgehend konstant bleiben. Die Ergebnisse zeigen nahezu ein Spiegelbild der beiden Korpora.

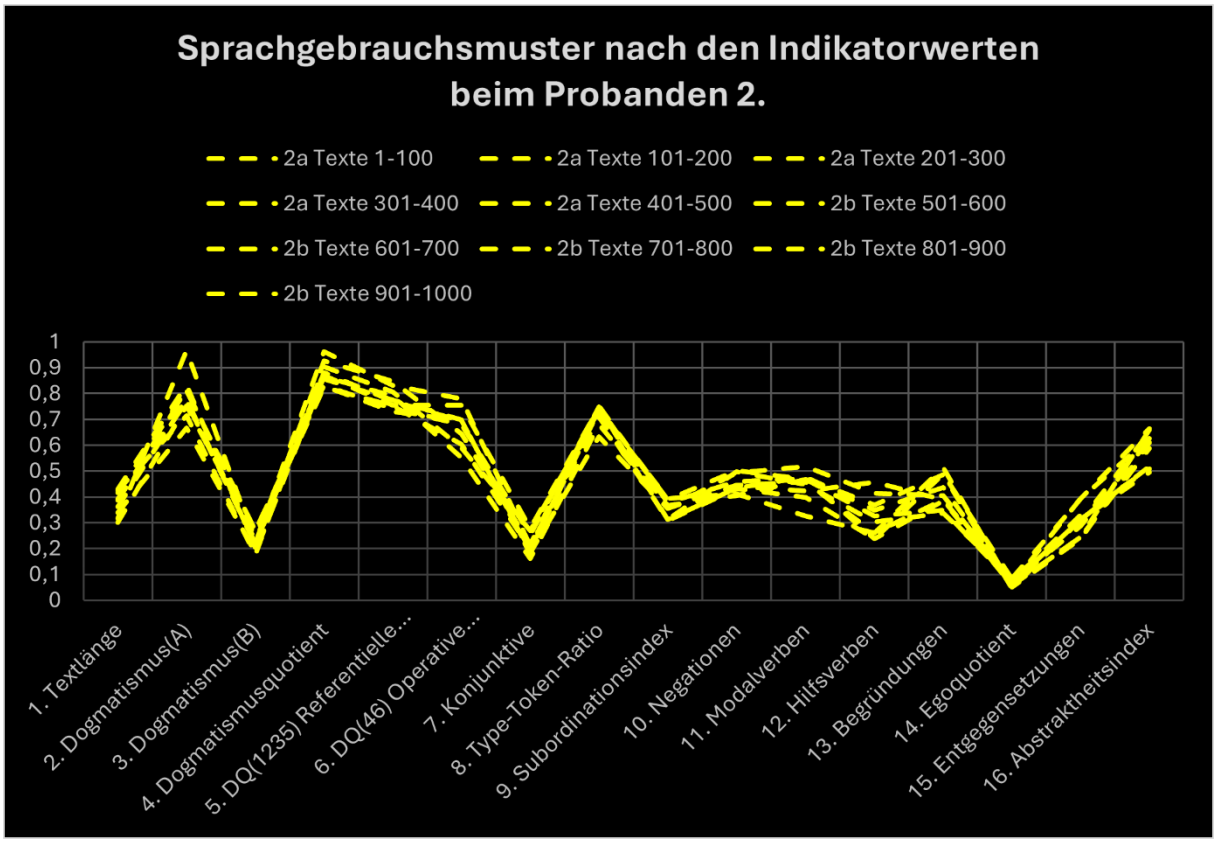
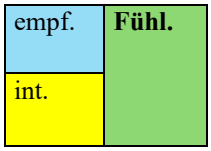


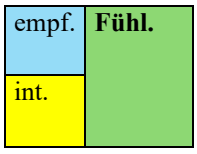
Abb. 10: Tendenzen der Sprachgebrauchsmuster beim Probanden 2

Basierend auf den ausgewählten Indikatorergebnissen lassen sich bei Probanden 2 die folgenden dominierenden Funktionen der menschlichen Informationsverarbeitung sowie mögliche Persönlichkeitsstile hypothetisieren. Bei ihm sind die Funktionen des Empfindens und Intuierens ausgeglichen. Die Fühlfunktion bleibt prädominant, obwohl die Denkfunktion nahezu gleich stark ausgeprägt ist. Die Ergebnisse lassen drei Annahmen zur möglichen Polysemie der Persönlichkeitsstile zu: Proband 2 könnte entweder eine Variante des selbstbestimmten, des ahnungsvollen oder des spontanen Persönlichkeitsstils aufweisen, wie in Abb. 11 dargestellt.

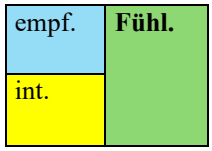
Proband 2	Empf.	Int.	Fühlen			
Indikator	Assoziierte Funktion		%	Indikator	Assoziierte Funktion	%
Subordinationsindex	IV (+) OE (-)		0.35	Abstraktheitsindex	AD (+) GF (-)	0.59
Textlänge	IV (+) OE (-)		0.37	Dogmatisches A	AD (-) GF (+)	0.78
Variationsindex	IV (+) OE (-)		0.72	Dogmatisches B	AD (+) GF (-)	0.22
Dominante Funktion	OE (52%) IV (48%)		OE/IV	Dominante Funktion	GF (66%) AD (44%)	GF



1. Selbstbestimmt a)



6. ahnungsvoll b)



11. spontan b)

Abb. 11: Die Ergebnisse der ausgewählten Indikatoren und die assoziierten Funktionen bzw. vermuteten Persönlichkeitsstile beim Probanden 2

4.3.4 Sprachgebrauchsmuster nach den Indikatorwerten beim Probanden 3

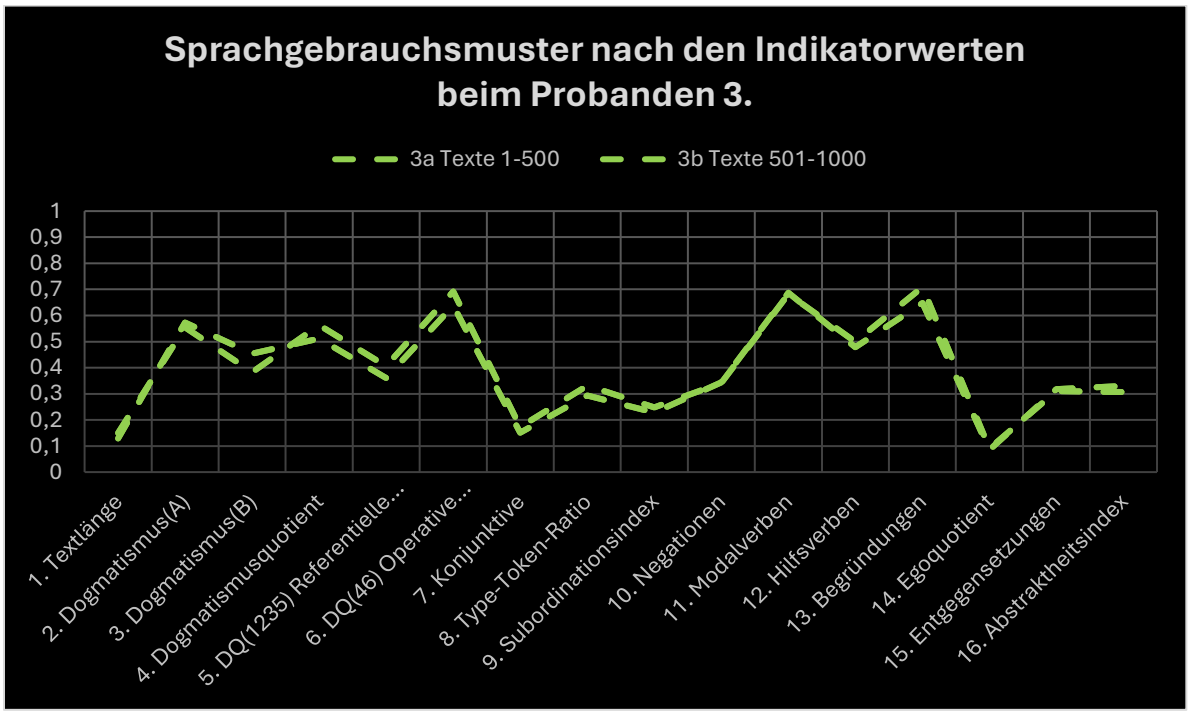


Abb. 12: Signifikante Übereinstimmungen der Sprachgebrauchsmuster beim Probanden 3

In Abb. 12 sind die Ergebnisse von Proband 3 gemäß der zuvor erläuterten Systematik dargestellt. Die Abb. zeigt deutlich, dass die Vergleichsgruppen 3a und 3b, bestehend aus den Texteinheiten 1–500 und 501–1000, keine signifikanten Abweichungen aufweisen. Abb. 13

bestätigt zudem die Vermutung der Überzufälligkeit der Sprachgebrauchsmuster, da die Tendenzen der Indikatorwerte relativ konstant sind. Die analysierten Korpora zeigen hierbei ebenfalls nahezu ein Spiegelbild.

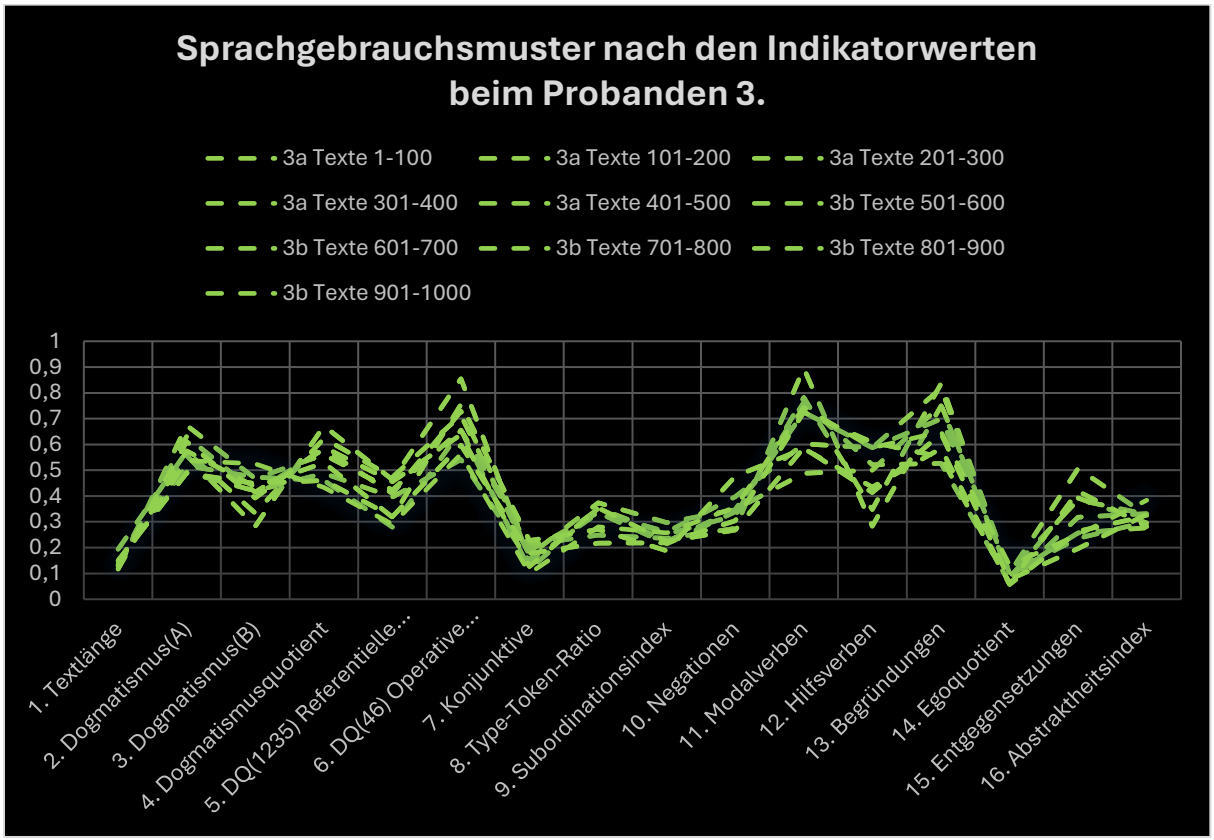
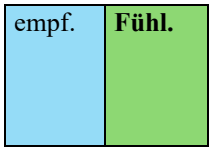


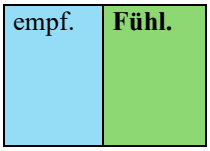
Abb. 13: Tendenzen der Sprachgebrauchsmuster beim Probanden 3

Nach den ausgewählten Indikatorergebnissen lassen sich bei Proband 3 die folgenden prädominierenden Funktionen der menschlichen Informationsverarbeitung hypothetisieren: Empfinden und Fühlen. Daraus ergibt sich eine Variante des selbstbestimmten oder sorgfältigen Persönlichkeitsstils, wie in Abb. 14 dargestellt.

Proband 3	Empfindendes	Fühlen			
Indikator	Assoziierte Funktion	%	Indikator	Assoziierte Funktion	%
Subordinationsindex	IV (+) OE (-)	0.23	Abstraktheitsindex	AD (+) GF (-)	0.31
Textlänge	IV (+) OE (-)	0.13	Dogmatisches A	AD (-) GF (+)	0.56
Variationsindex	IV (+) OE (-)	0.31	Dogmatisches B	AD (+) GF (-)	0.41
Dominante Funktion	OE (88%) IV (22%)	OE	Dominante Funktion	GF (62%) AD (38%)	GF



1. Selbstbestimmt b)



5. sorgfältig b)

Abb. 14: Die Ergebnisse der ausgewählten Indikatoren und die assoziierten Funktionen bzw. vermuteten Persönlichkeitsstile beim Probanden 3

4.3.5 Sprachgebrauchsmuster nach den Indikatorwerten beim Probanden 4

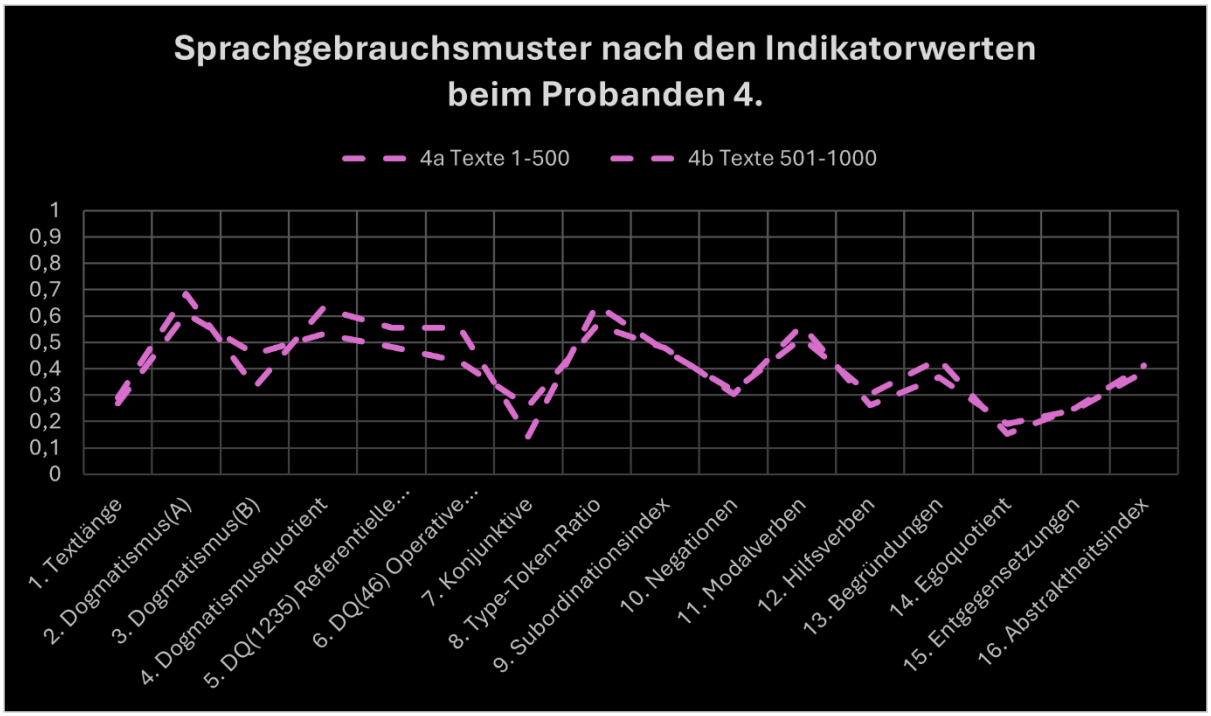


Abb. 15: Signifikante Übereinstimmungen der Sprachgebrauchsmuster beim Probanden 4

In Abb. 15 sind die Ergebnisse des Probanden 4 gemäß der zuvor beschriebenen Systematik dargestellt. Die Abb. zeigt deutlich, dass die Vergleichsgruppen 4a und 4b mit den Texteinheiten 1–500 und 501–1000 keine signifikanten Abweichungen aufweisen.

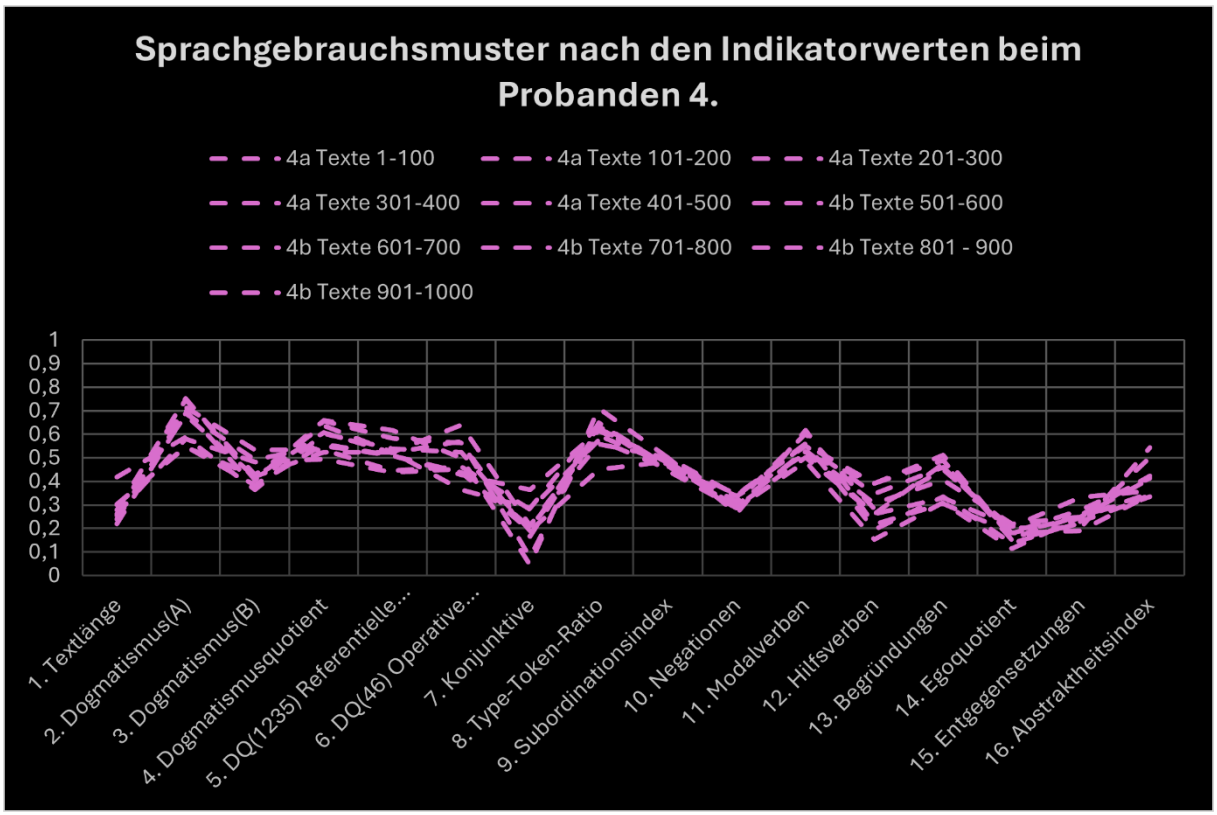
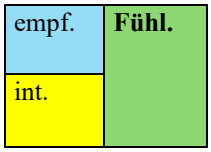


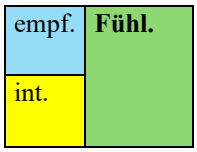
Abb. 16: Tendenzen der Sprachgebrauchsmuster bei dem Probanden 4

Abb. 16 untermauert zudem die Vermutung der Überzufälligkeit der Sprachgebrauchsmuster, da die Tendenzen der Indikatorwerte weitgehend konstant bleiben. Basierend auf den ausgewählten Indikatorergebnissen lassen sich bei Proband 4 ausgeglichene Funktionen des Empfindens und Intuierens sowie eine prädominante Funktion des Fühlens vermuten. In Abb. 17 sind die möglichen Persönlichkeitsstile dargestellt.

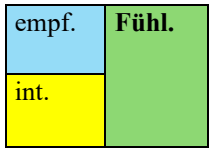
Proband 4	Empf.	Int.	Fühlen			
Indikator	Assoziierte Funktion		%	Indikator	Assoziierte Funktion	%
Subordinationsindex	IV (+) OE (-)		0.47	Abstraktheitsindex	AD (+) GF (-)	0.40
Textlänge	IV (+) OE (-)		0.27	Dogmatisches A	AD (-) GF (+)	0.64
Variationsindex	IV (+) OE (-)		0.60	Dogmatisches B	AD (+) GF (-)	0.42
Dominante Funktion	OE (56%) IV (44%)		OE/IV	Dominante Funktion	GF (61%) AD (39%)	GF



1. Selbstbestimmt a)



6. ahnungsvoll b)



11. spontan b)

Abb. 17: Die Ergebnisse der ausgewählten Indikatoren und die assoziierten Funktionen bzw. vermuteten Persönlichkeitsstile beim Probanden 4

4.3.6 Sprachgebrauchsmuster nach den Indikatorwerten beim Probanden 5

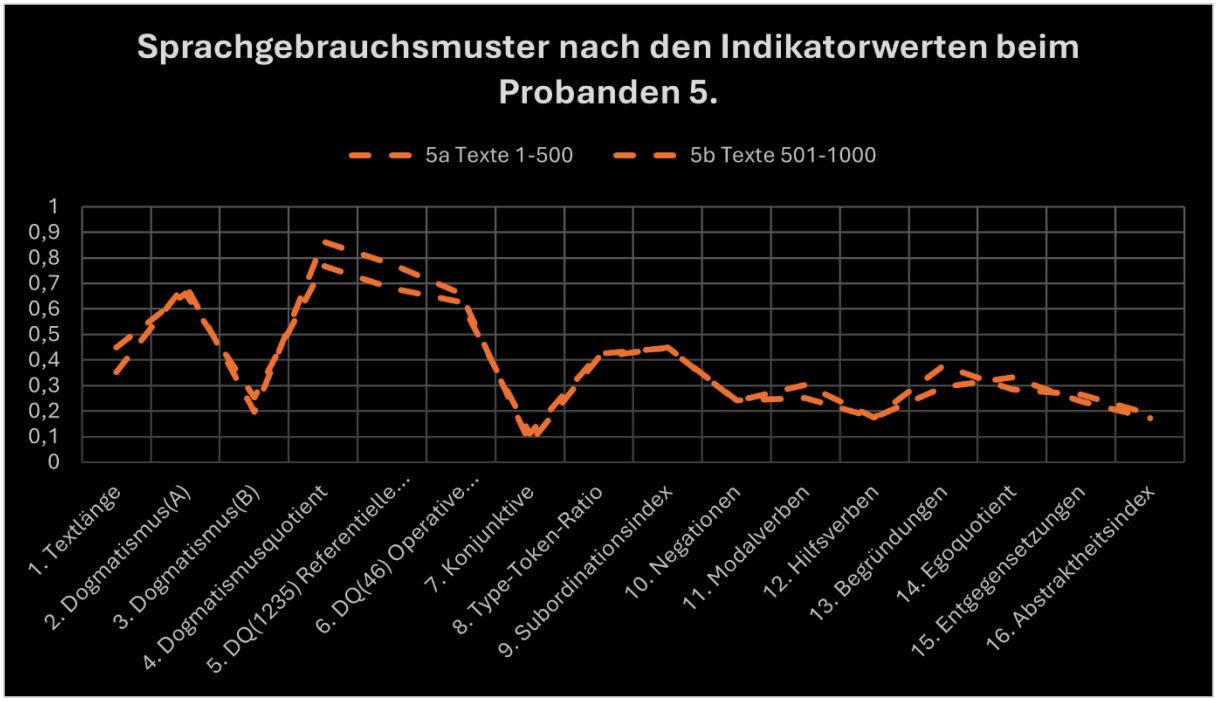


Abb. 18: Signifikante Übereinstimmungen der Sprachgebrauchsmuster beim Probanden 5

In Abb. 18 sind die Ergebnisse von Proband 5 nach der zuvor beschriebenen Systematik dargestellt. Die Abb. zeigt deutlich, dass die Vergleichsgruppen 5a und 5b, bestehend aus den Texteinheiten 1–500 bzw. 501–1000, keine signifikanten Abweichungen aufweisen. Abb. 19

untermauert zudem die Annahme der Überzufälligkeit der Sprachgebrauchsmuster, da die Tendenzen der Indikatorwerte relativ konstant bleiben.

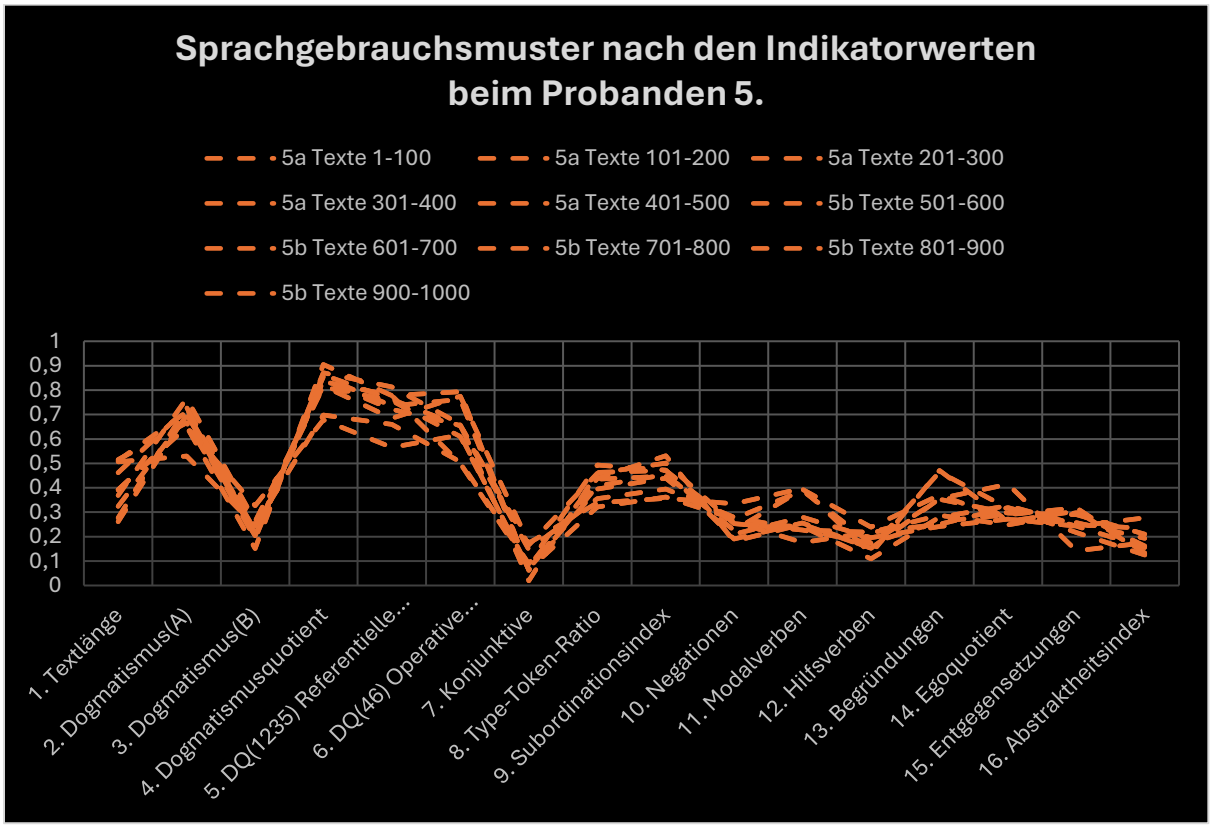
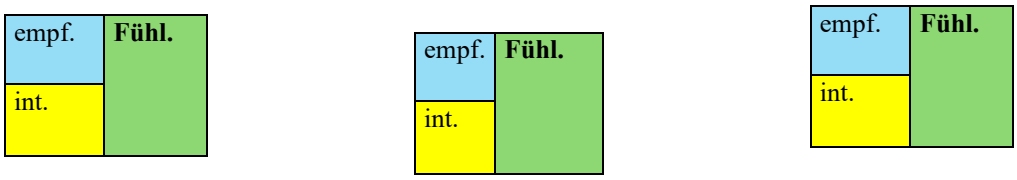


Abb. 19: Tendenzen der Sprachgebrauchsmuster beim Probanden 5

Auf Grundlage der ausgewählten Indikatorergebnisse lassen sich bei Proband 5 die folgenden dominierenden Funktionen der menschlichen Informationsverarbeitung sowie die möglichen Persönlichkeitsstile hypothetisieren, wie in Abb. 20 dargestellt.

Proband 5	Empf.	Int.	Fühlen			
Indikator	Assoziierte Funktion		%	Indikator	Assoziierte Funktion	%
Subordinationsindex	IV (+) OE (-)		0.44	Abstraktheitsindex	AD (+) GF (-)	0.18
Textlänge	IV (+) OE (-)		0.40	Dogmatisches A	AD (-) GF (+)	0.67
Variationsindex	IV (+) OE (-)		0.41	Dogmatisches B	AD (+) GF (-)	0.22
Dominante Funktion	OE (59%) IV (41%)		OE/IV	Dominante Funktion	GF (76%) AD (24%)	GF



1. Selbstbestimmt a) *6. ahnungsvoll b)* *11. spontan b)*

Abb. 20: Die Ergebnisse der ausgewählten Indikatoren und die assoziierten Funktionen bzw. vermuteten Persönlichkeitsstile beim Probanden 5

4.3.7 Vergleich der Indikatorwerte bei den Probanden 1–5

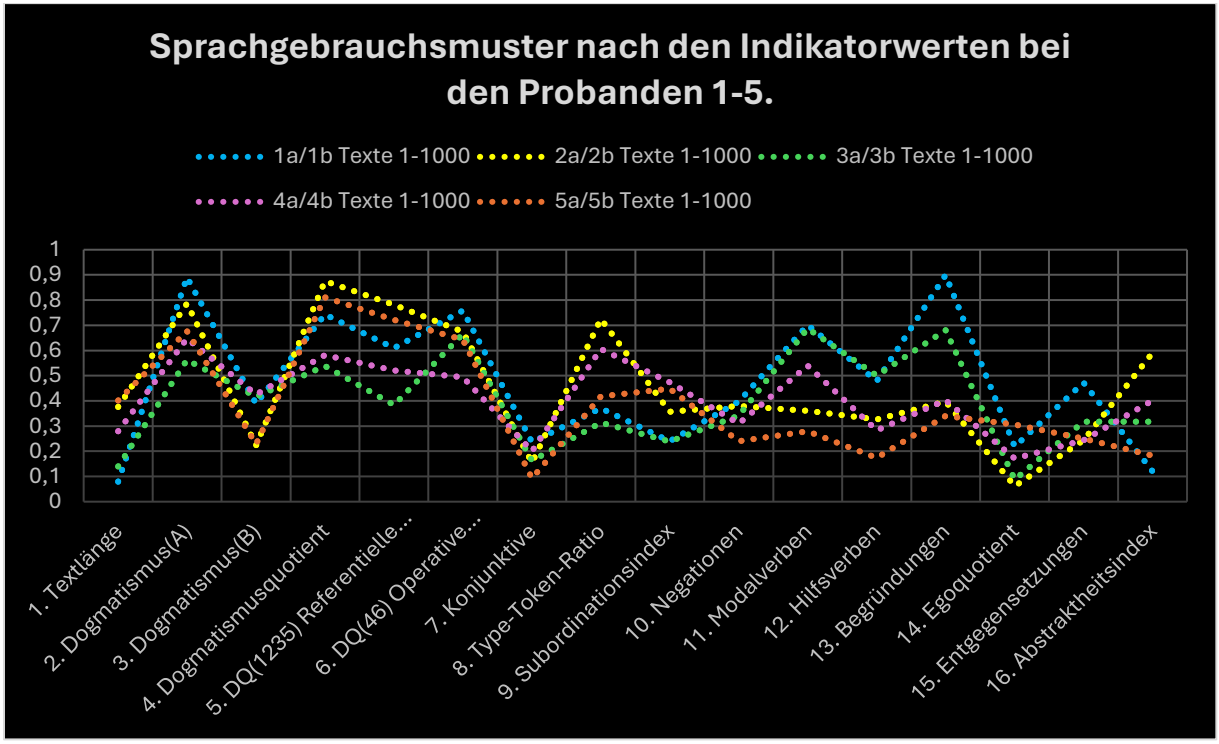


Abb. 21: Signifikante Abweichungen der Sprachgebrauchsmuster bei den Probanden 1–5

In Abb. 21 sind die Ergebnisse der Probanden 1–5 mit jeweils 1000 Texteinheiten dargestellt. Die Abb. fasst die Ergebnisse dahingehend zusammen, dass die Sprachgebrauchsmuster der Probanden im Verhältnis zueinander in mehreren Punkten eine statistisch signifikante Abweichung aufweisen – wie es in der Darstellung der Z-Testergebnisse detailliert erläutert

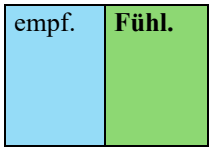
wird. Während die letzten zehn Abbildungen zeigen, dass es innerhalb der Vergleichsgruppen eines gegebenen Probanden keine statistisch signifikante Abweichung gibt, verdeutlicht Abb. 21, dass zwischen den Indikatorwerten der einzelnen Probanden in mehreren Punkten eine statistisch signifikante Abweichung besteht. Die Ergebnisse bestätigen sowohl die Hypothese über die Wertekonstanz der Sprachgebrauchsmuster als auch die Hypothese über die Diskrepanzwerte zwischen den Sprachbenutzern.

Die Abweichungen und Übereinstimmungen der Werte zwischen den Probanden lassen sich dadurch erklären, dass sich die Informationsverarbeitungsstile der Probanden 1–5 tendenziell den Durchschnittswerten annähern. Drei Probanden waren vorwiegend empfindend-fühlend (Probanden 1 und 3), während die anderen drei Probanden eine ausgewogene empfindend/intuierend-fühlende Verarbeitungsweise aufwiesen (Probanden 2, 4 und 5). Diese Verarbeitungsstile können als ähnlich betrachtet werden. Trotz ihrer Ähnlichkeit unterscheiden sie sich jedoch in ihrer spezifischen Zusammensetzung – sie sind kontingent zusammengesetzt. Kurz gesagt: Die Übereinstimmungen der Werte lassen sich darauf zurückführen, dass die zugrunde liegenden Informationsverarbeitungsstile ebenfalls übereinstimmen. Die Unterschiede hingegen ergeben sich aus der spezifischen Zusammensetzung der dominierenden Funktionen innerhalb der jeweiligen Persönlichkeitsstile. Obwohl die Werte persönlichkeitsspezifische Muster erkennen lassen, bleiben die Werte der einzelnen Teilsysteme stark personenspezifisch, sind in ihrer Verteilung jedoch unabhängig. Selbst wenn eine Person demselben Persönlichkeitsstil zugeordnet wird, zeigen sich dennoch individuelle Variationen.

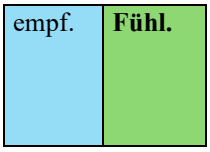
4.3.8 Die Durchschnittswerte bei den Probanden 1–55

Basierend auf den ausgewählten Indikatorergebnissen lassen sich für die Probanden 1–55 die folgenden Durchschnittswerte der dominierenden Funktionen der menschlichen Informationsverarbeitung sowie die häufigsten Persönlichkeitsstile hypothetisieren. Der Großteil der zufällig ausgewählten Probanden weist einen Informationsverarbeitungsstil des empfindenden Fühlens auf, wobei sie entweder einer Variante des selbstbestimmten oder des sorgfältigen Persönlichkeitsstils zugeordnet werden können, wie in Abb. 22 ersichtlich ist.

Durchschnitt	Empfindendes	Fühlen			
Indikator	Assoziierte Funktion	%	Indikator	Assoziierte Funktion	%
Subordinationsindex	IV (+) OE (-)	0.33	Abstraktheitsindex	AD (+) GF (-)	0.29
Textlänge	IV (+) OE (-)	0.18	Dogmatisches A	AD (-) GF (+)	0.62
Variationsindex	IV (+) OE (-)	0.53	Dogmatisches B	AD (+) GF (-)	0.35
Dominante Funktion	OE (66%) IV (34%)	OE	Dominante Funktion	GF (66%) AD (34%)	GF



1. Selbstbestimmt b)



5. sorgfältig b)

Abb. 22: Die Ergebnisse der ausgewählten Indikatoren sowie der assoziierten Funktionen und vermuteten Persönlichkeitsstile auf Basis der Durchschnittswerte der Probanden 1–55

Basierend auf den ausgewählten Indikatorergebnissen lassen sich die dominierenden Funktionen bei den Probanden 1–55 in der folgenden Häufigkeit vermuten. Falls die Ergebnisse als repräsentativ betrachtet werden können, lässt sich die Annahme aufstellen, dass in der Mehrheit der Population ein fühlender bzw. empfindender Stil vorherrscht. Die Kookkurrenzen des empfindenden Fühlens machen nahezu die Hälfte der identifizierten Persönlichkeitsstile aus. Intuieren und Denken treten hingegen seltener auf. Die in der Popkultur bekannten Typen des intuitiven Denkens sind unter den Probanden überhaupt nicht vertreten. Abb. 23 zeigt die Häufigkeit der assoziierten Informationsverarbeitungsstile.

OES/IVS	GF/AD	Häufigkeit
	Fühlen	74.5%
Empfindendes		66.4%
	Denken	5.4%
Intuitives		5.4%
	Denk. Fühl.	23.6%
Empf. Int.		29.0%

OES/IVS	GF/AD	Häufigkeit
Intuitives	Denken	0%
Intuitives	Fühlen	1.8%
Empfindendes	Denken	3.6%
Empfindendes	Fühlen	45%
Empf. Int.	Denken	1.8%
Empf. Int.	Fühlen	20%
Empf. Int.	Denk. Fühl.	7.2%
Empfindendes	Denk. Fühl.	10.9%
Intuitives	Denk. Fühl.	3.6%

Abb. 23: Häufigkeit der assoziierten Funktionen der Informationsverarbeitungsstile bei den Probanden 1–55

Die Ergebnisse ähneln den Anteilen in computergestützten Untersuchungen mit ähnlichen, auf Jung basierenden Systemen. Eine Studie von Makwana und Govind, die MBTI-Variablen

quantifizierte, fand unter Managementstudenten einen Anteil von 68,5 % für Empfinden, 29,5 % für Intuieren, 63,6 % für Fühlen und 40,9 % für Denken (vgl. Makwana/Govind 2020: 31). Es ist allerdings festzustellen, dass die MBTI-Variablen mit binären Gegensätzen operieren und keinen ausgewogenen Sowohl-als-auch-Zustand kennen.

4.4 Darstellung der Z-Test-Ergebnisse

Die Ergebnisse der Z-Tests bestätigen die beiden Hypothesen mit den jeweiligen statistisch signifikanten Abweichungen und Übereinstimmungen. Die Daten der 16 sprachstatistischen Indikatoren waren die folgenden: Wortzahl, dogmatische A-Ausdrücke, dogmatische B-Ausdrücke, Dogmatismusquotient, referentielle Prägnanz, operative Prägnanz, Konjunktionen, Variationsindex, Subordinationsindex, Negationen, Modalverben, Hilfsverben, Begründungen, Egoquotient, Entgegensetzungen und Abstraktheitsquotient. Andererseits ging es um die jeweiligen Deviationsdaten zwischen den Vergleichsgruppen, die Durchschnittswerte (μ) und die Abweichungswerte (σ), nach denen die kritischen Z-Werte berechnet wurden. Die Standardabweichung wurde anhand der Abweichungswerte der Indikatorwerte zwischen den jeweiligen a- und b-Vergleichsgruppen berechnet. Zu jedem Indikatorwert gehören dementsprechend fünf Abweichungswerte, zu denen die weiteren Abweichungen ins Verhältnis gesetzt wurden. Das μ gibt dementsprechend stets den Durchschnitt der jeweils fünf Abweichungswerte zwischen den a- und b-Vergleichsgruppen an, nach dem alle Abweichungen parametrisiert bzw. verglichen wurden. Die jeweiligen μ-, σ- und Z-Werte wurden mit Python automatisiert berechnet, basierend auf den folgenden Formeln:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Formel 17: Die statistische Berechnung der μ-Werte (Durchschnitt der Abweichungswerte), nach der die Berechnung mit Python automatisiert wurde (vgl. Aron/Coups/Aron 2013: 2)

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{n}}$$

Formel 18: Die statistische Berechnung der σ-Werte (Streuung der Abweichungswerte), nach der die Berechnung mit Python automatisiert wurde (vgl. Aron/Coups/Aron 2013: 2)

$$z = \frac{x - \mu}{\sigma}$$

Formel 19: Die Formel zur statistischen Berechnung der z-Werte (das statistische Signifikanzkriterium der Abweichungen), nach der die Berechnung mit Python automatisiert wurde (vgl. Aron/Coups/Aron 2013: 2)

Der kritische Z-Wert liegt bei etwa 2,575. Überschreitet der berechnete Z-Wert in den einzelnen Z-Tests diesen kritischen Wert, kann von einer signifikanten Abweichung gesprochen werden.

Bleibt der berechnete Z-Wert hingegen unter dem Niveau des kritischen Z-Wertes, lässt sich keine signifikante Abweichung feststellen. Im Mittelpunkt der Origo der Signifikanztabellen liegt stets eine Vergleichsgruppe, zu der die anderen Angaben relativiert wurden.

5. Zusammenfassung und Ausblick

Die vorliegende Arbeit konzentriert sich auf die Beantwortung zweier Hypothesen. Mithilfe der 16 sprachstatistischen Indikatoren lässt sich eine sprachliche Profilierung der Persönlichkeitsmerkmale der jeweiligen Sprachbenutzer durchführen, sofern eine ausreichende Qualität und Quantität sprachlicher Materialien für die Korpusbildung vorliegt. Die Sprachgebrauchsmuster erweisen sich als überzufällig. Bereits mit den 16 Indikatoren können die Sprachgebrauchswerte effektiv voneinander abgegrenzt werden, auch ohne ein Gruppierbarkeitskriterium. Die sechs ausgewählten Indikatoren, auf deren Grundlage die möglichen Persönlichkeitsstile berechnet wurden, liefern ein Gesamtbild über die potenziellen Tendenzen. Die Entwicklung weiterer Indikatoren könnte zukünftige Untersuchungen zweifellos noch effektiver gestalten.

Die angewendete Methode folgte der Empfehlung Müllers (2013: 21): „Wenn komplexe Zusammenhänge gedanklich nachvollzogen werden sollen, hilft zunächst eine drastische Vereinfachung des zu erklärenden Systems“. Ob es sich um kognitiv-emotionale Allomorphe handelt (vgl. Bühler et al. 2010: 487), die sich auf der Sprachoberfläche manifestieren und dabei mehrdeutig sind, soll für zukünftige Forschungen mit komplexem Data-Science-Design eine fruchtbare Grundlage bieten. Obwohl die Indikatorwerte des Sprachgebrauchs relativ konstant sind, kann über ihre exakte Bedeutung nur das neuronale Funktionieren des jeweiligen Sprachbenutzers Aufschluss geben. Ohne diese Erkenntnisse lassen sich grundlegende Dilemmata nicht lösen – etwa die Frage, ob ein abstraktes Wort tatsächlich durch eine abstrakte neuronale Repräsentation ermöglicht wird oder ob es lediglich abstrakt erscheint. Die Kartierung neuronaler Informationen hinter kognitiv-emotionalen Allomorphen könnte wertvolle Einblicke in die Natur bedeutungstragender Elemente liefern, die die menschliche Gestaltwahrnehmung prägen. Es ist möglich, dass die Sprachgebrauchswerte konstant zu sein scheinen, während ihre Bedeutung bereits polysem ist. Daher liegt die Vermutung nahe, dass es sich um kognitiv-emotionale Bedeutungsträger handelt, die die jeweilige Neurokomplexität widerspiegeln – Allomorphe bzw. Alloseme (vgl. Hoffmann 2010: 465) die neurospezifisch vernetzt sind. Die Interpretation sprachlichen Verhaltens könnte ebenso mehrdeutig sein wie die Deutung menschlichen Verhaltens im Allgemeinen.

Es wirft schwerwiegende Fragen auf, in welchem Maße die sprachstatistischen Ergebnisse in unterschiedlichen Deutungsrahmen bzw. Persönlichkeitskonzepten mehrdeutig erscheinen können. Bedeutet ein überdurchschnittlicher Wert der Textlänge eine erhöhte Offenheit oder eine ausgeprägte Extraversion? Deuten A-Ausdrücke eher auf Gewissenhaftigkeit oder auf Neurotizismus hin? Lassen sich hinter Negationswörtern eher Merkmale von Gewissenhaftigkeit oder von Verträglichkeit vermuten? Die einzig treffende Antwort wäre: Es kommt darauf an. Es ist Aufgabe zukünftiger Korrelationsuntersuchungen, diese Fragen zu klären. Durch ein Big-Data-Setting mit einem ähnlichen Untersuchungsgegenstand könnte sich die Wahrscheinlichkeit äußerst präzise bestimmen lassen (vgl. Müller 2013: 16), inwieweit sich von der reinen Sprachverwendung auf statistisch erfasste Neurokomplexität schließen oder diese aus Big-Data-Modellen ableiten lässt. All dies wirft jedoch schwerwiegende ethische Fragen auf – insbesondere, wenn man bedenkt, dass eine erfolgreiche sprachliche Profilierung der menschlichen Persönlichkeit bereits durch die Methode primitiver Verteilungswerte möglich wurde.

Obwohl die verwendete Methode auf eine drastische Vereinfachung des Analysegegenstandes beruht, war es bereits möglich, einen Einblick in die Natur der Kohärenzmuster des individuellen Sprachgebrauchs zu gewinnen. Eine mögliche Weiterentwicklung der Arbeit könnte in der qualitativen diskurstheoretischen Untersuchung des Analysegegenstandes bestehen. Nach Klug (2017: 76) können wir zwischen denotativem und deontischem Wissen unterscheiden. Während die Kategorie des denotativen Wissens die objektive Struktur der wahrnehmbaren Wirklichkeit darstellt, erhält das deontische Wissen eine persönlichkeitsspezifische Färbung, da die menschliche Persönlichkeit als Zustand der Strukturiertheit der Wissensbestände bzw. der Beziehungen zwischen zNS (zentrales Nervensystem) und kNS (konzeptuelles Nervensystem) modelliert werden kann (vgl. Schubert 2008: 14). Es könnte die Aufgabe einer qualitativen Analyse sein, die Art und Weise zu ermitteln, wie Informationsverarbeitungsstile der menschlichen Persönlichkeit hinter kognitiv-emotionalen Allomorphen und Allosemen ihre jeweiligen kognitiven Färbungen an die Struktur der wahrgenommenen Wirklichkeit binden (vgl. Andresen 2015: 264). Wittgensteins Satz, „Die Grenzen meiner Sprache bedeuten die Grenzen meiner Welt“, erscheint im Sinne einer kognitiven Deutung des neuronalen Funktionierens in einem neuen Licht (vgl. Müller 2013: 183).

Literaturverzeichnis

Andresen, Liv (2015): Persönlichkeitsspezifische Sprachvariation: Zum sprachwissenschaftlichen Potential psychologischer Methoden. In: Zeitschrift für Dialektologie und Linguistik 3, 263–285. https://doi.org/10.25162/zdl-2015-0009

Aron, Arthur/Coups, Eliot/Aron, Elanie (2013): Statistics for Psychology. 6. Aufl. New York, London, Munich: Pearson.

Barsalou, Lawrence (2012): The Human Conceptual System. In: Spivey, Michael J./McRae, Ken/Joanisse, Marc F. (Hg.): The Cambridge Handbook of Psycholinguistics. Cambridge, New York: Cambridge University Press, 239–258. https://doi.org/10.1017/cbo9781139029377.013

Bock, Herbert (2021): Symposium Human Communication: „Kommunikation oder Trennukation“, https://www.youtube.com/watch?v=qNZ4r5L9bTs.

Brown, Penelope/Levinson, Stephen C. (1987): Politeness. Some Universals in Language Usage. Cambridge, New York: Cambridge University Press. https://doi.org/10.1017/cbo9780511813085

Bühler, Hans et al. (2010): Linguistik I: Einführung in die Morphemik (1970). In: Hoffmann, Ludger (Hg.): Sprachwissenschaft. Ein Reader. 3., aktualisierte und erweiterte Aufl. Berlin, New York: De Gruyter, 478–496. https://doi.org/10.1515/9783110226300.5.478

Chomsky, Noam (1995): Looking for the Magic Answer? Noam Chomsky interviewed by David Barsamian. In: https://chomsky.info/warfare03

Chomsky, Noam (2010): Probleme sprachlichen Wissens: I. Ein Rahmen für die Diskussion (1988). In: Hoffmann, Ludger (Hg.): Sprachwissenschaft. Ein Reader. 3., aktualisierte und erweiterte Aufl. Berlin, New York: De Gruyter, 114–129. https://doi.org/10.1515/9783110226300.1.114

Dell, Gary S./Cholin, Joana (2012): Language Production. Computational Models. In: Spivey, Michael J./McRae, Ken/Joanisse, Marc F. (Hg.): The Cambridge Handbook of Psycholinguistics. Cambridge, New York: Cambridge University Press, 426–442. https://doi.org/10.1017/cbo9781139029377.022

Dietrich, Rainer/Gerwien, Johannes (2017): Psycholinguistik. Eine Einführung. 3. Aufl. Stuttgart: Metzler. https://doi.org/10.1007/978-3-476-05494-4

Diekmann, Andreas (2011): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek bei Hamburg: Rowohlt.

Graesser, Arthur C./MacNamara, Danielle S./Rus, Vasile (2012): Computational Modeling of Discourse and Conversation. In: Spivey, Michael J./McRae, Ken/Joanisse, Marc F. (Hg.):

The Cambridge Handbook of Psycholinguistics. Cambridge, New York: Cambridge University Press, 558–572. https://doi.org/10.1017/cbo9781139029377.029

Griffin, Zensi/Crew, Christopher (2012): Research in Language Production. In: Spivey, Michael J./McRae, Ken/Joanisse, Marc F. (Hg.): The Cambridge Handbook of Psycholinguistics. Cambridge, New York: Cambridge University Press, 409–425. https://doi.org/10.1017/cbo9781139029377.021

Hoffmann, Svenja (2023): *Diagnose-Revolution? Smartphone kann Diabetes nach wenigen Sekunden erkennen!* 19.10.2023. *rtl.de*. Online unter: https://www.rtl.de/cms/forschung-smartphone-erkennt-typ-2-diabetes-mediziner-ordnet-ein-5063472.html.

Jung, Carl (2017): Psychological Types. New York, London: Routledge. https://doi.org/10.4324/9781315512334

Klug, Maria (2017): Multimodale Text- und Diskursanalyse. In: Der Deutschunterricht 6, 73–85.

Kuhl, Julius/Quirin, Markus/Koole, Sander L. (2020): The Functional Architecture of Human Motivation: Personality Systems Interactions Theory. In: Elliot, Andrew (Hg.): Advances in Motivation Science. Bd. 7, 1–62. https://doi.org/10.1016/bs.adms.2020.06.001

Lakoff, George/Wehling, Elisabeth (2010): Auf leisen Sohlen ins Gehirn: Im Land der zwei Freiheiten: Warum wir hören, was wir denken (2008). In: Hoffmann, Ludger (Hg.): Sprachwissenschaft. Ein Reader. 3., aktualisierte und erweiterte Aufl. Berlin, New York: De Gruyter, 147–154. https://doi.org/10.1515/9783110226300.1.147

Makwana Kirti/Govind, Dave (2020): A Study of Identification of Personality Profiles of Undergraduate Management Students Using Myers Briggs Type Indicator (MBTI) Test. In: Pacific Business Review International 8, 1–31.

Müller, Horst M. (2013): Psycholinguistik – Neurolinguistik. Die Verarbeitung von Sprache im Gehirn. Paderborn: Fink. https://doi.org/10.36198/9783838536477

Oakes, Michael (2022): Author Profiling and Related Applications. In: Mitkov, Ruslan (Hg.): The Oxford Handbook of Computational Linguistics. 2. Aufl. Oxford: Oxford University Press, 1165–1197. https://doi.org/10.1093/oxfordhb/9780199573691.013.53

Pichler, Georg/Pramer, Philip (2023): Warum 2023 das Jahr der künstlichen Intelligenz wird. In: Der Standard, 15.1.2023, https://www.derstandard.de/story/200014257 0218/warum-2023-das-jahr-der-kuenstlichen-intelligenz wird

Safran, Jenny R./Sahni, Sarah D. (2012): Learning the Sounds of Language. In: Spivey, Michael/McRae, Ken/Joanisse, Marc: The Cambridge Handbook of Psycholinguistics.

Cambridge, New York: Cambridge University Press, 42–60. https://doi.org/10.1017/cbo9781139029377.004

Schubert, Franziska (2008): Sprache und Persönlichkeit. Differentielles Ausdrucksverhalten unter Berücksichtigung der Sprachsituation. Dresden, Diss.

Zenthöfer, Jochen (2023): Erste Uni schafft Bachelorarbeiten ab. In: Frankfurter Allgemeine Zeitung, 13.01.2024, https://www.faz.net/aktuell/karriere-hochschule/hoersaal/ki-und-plagiate-erste-uni-schafft-bachelorarbeiten-ab-19353621.html