

Eine neue Perspektive auf
Berufswahltheorien:
Wie Künstliche Intelligenz
Interessensstrukturmodelle anhand
menschlicher Sprache versteht

MASTER-ARBEIT

2025

Autor

Hêche, Nicolas

Begleitpersonen

Theiler, Sven

Prof. Dr. Sterchi, Yanik

Praxispartner

Fachhochschule Nordwestschweiz FHNW
Hochschule für Angewandte Psychologie APS

Prof. Dr. Hell, Benedikt

Abstract

An important milestone in every person's life is choosing a fitting profession, which is largely influenced by individual interests. The field of vocational interests addresses how a suitable profession can be found. One of its most established theories is Holland's theory, also known as the RIASEC model. Based on it, numerous instruments for assessing vocational interests have been developed. One is the Verb Interest Test (VIT), which assesses them using decontextualized Verbs. Each verb is assigned to one of the six RIASEC dimensions and represents it. As artificial intelligence (AI) advances into various domains, the question arises how it might be applied in the context of vocational interests. Word embedding, originating from AI, enables not only the representation of word meanings in a vector space, but also the discovery of latent semantic structures. This study investigates to what extent the structure of the RIASEC model is reflected in a word embedding. To answer this, the VIT's verbs were analyzed in a pre-trained word embedding. Using principal component analysis, six factors were extracted from the verb vectors. These were analyzed based on the mean factor loadings of the verbs for each RIASEC dimension as well as the verbs with the ten strongest positive and negative factor loadings. None of the extracted factors clearly reflected a single model dimension. The first two factors were the most interpretable, while the remaining represented mixed factors. Overall, the results suggest a lower degree of discriminability between RIASEC dimensions as well as more complex, blended interest spaces.

Key words: artificial intelligence (AI), natural language processing (NLP), word embedding, vocational interest, RIASEC, Verb-Interest-Test (VIT), principal component analysis

Zusammenfassung

Ein wichtiger Meilenstein im Leben jedes Menschen ist die Wahl eines geeigneten Berufes, die massgeblich durch individuelle Interessen beeinflusst wird. Das Forschungsfeld der Berufsinteressen beschäftigt sich mit der Frage, wie ein passender Beruf gefunden werden kann. Eine der etabliertesten Theorien dieses Bereichs ist Hollands Theorie, auch bekannt als RIASEC-Modell. Basierend auf dieser wurden zahlreiche Messinstrumente zur Erfassung von Berufsinteressen entwickelt. Eines davon ist der Verb-Interessenstest (VIT), welcher Berufsinteressen mithilfe kontextloser Tätigkeitsverben erfasst. Diese Verben sind jeweils einer der sechs RIASEC-Dimensionen zugeordnet und repräsentieren diese. Gleichzeitig stellt sich durch das Voranschreiten künstlicher Intelligenz (KI) in verschiedenste Bereiche die Frage nach deren Anwendung im Kontext der Berufsinteressen. Das Verfahren der Word-Embeddings aus dem Bereich der KI erlaubt nicht nur die Abbildung von Wortbedeutungen in einem Vektorraum, sondern auch das Entdecken latenter Bedeutungsstrukturen. Ziel dieser Arbeit ist es zu untersuchen, inwiefern sich die Struktur des RIASEC-Modells in einem Word-Embedding widerspiegelt. Zur Beantwortung dieser Frage wurden die Verben des VITs in einem vortrainierten Word-Embedding analysiert. Mittels Hauptkomponentenanalyse wurden aus den Vektoren der Verben sechs Faktoren extrahiert. Die Analyse der Faktoren erfolgte anhand der Mittelwerte der Faktorladungen der Verben pro RIASEC-Dimension, sowie der Verben mit den zehn stärksten positiven und negativen Faktorladungen. Keiner der extrahierten Faktoren spiegelte dabei eine Dimension des Modells eindeutig wider. Von den sechs Faktoren konnten die ersten beiden am deutlichsten interpretiert werden, während die restlichen Mischfaktoren darstellten. Die Ergebnisse deuten insgesamt auf eine geringere Trennschärfe zwischen den RIASEC-Dimensionen sowie auf komplexere, gemischte Interessensräume hin.

Schlagwörter: Künstliche Intelligenz (KI), Natural Language Processing (NLP) Word-Embedding, Berufsinteressen, RIASEC-Modell, Verb-Interessenstest (VIT), Hauptkomponentenanalyse

Inhaltsverzeichnis

| | |
|---|----|
| Abstract | 2 |
| Zusammenfassung | 3 |
| 1 Einleitung | 7 |
| 1.1 Ausgangslage | 7 |
| 1.2 Zielsetzung | 9 |
| 1.3 Aufbau der Arbeit..... | 10 |
| 2 Theoretische Grundlagen | 10 |
| 2.1 Künstliche Intelligenz und Sprachverarbeitung | 10 |
| 2.1.1 Künstliche Intelligenz | 11 |
| 2.1.2 Natural Language Processing..... | 14 |
| 2.2 Word-Embeddings | 15 |
| 2.2.1 Grundlagen von Word-Embeddings..... | 16 |
| 2.2.2 Unterschiedliche Word-Embedding Modelle..... | 19 |
| 2.3 Berufsbezogene Interessen..... | 22 |
| 2.3.1 Interessen..... | 23 |
| 2.3.2 Hollands Theorie (RIASEC) | 25 |
| 2.3.3 Erweiterungen des RIASEC-Modells | 29 |
| 2.3.4 Verbe-Interessenstest | 31 |
| 3 Methodisches Vorgehen | 38 |
| 3.1 Auswahl eines Word-Embeddings | 38 |
| 3.1.1 Abgleich von Word-Embedding und VIT..... | 40 |
| 3.1.2 Intrinsische Evaluation | 40 |
| 3.2 Vorbereitung der Daten für die Hauptkomponentenanalyse..... | 42 |

| | | |
|---|--|----|
| | | 5 |
| | 3.3 Hauptkomponentenanalyse | 43 |
| 4 | Ergebnisse | 47 |
| | 4.1 Ergebnisse der Auswahl eines Word-Embeddings | 47 |
| | 4.1.1 Ergebnisse des Abgleichs von VIT und Word-Embeddings..... | 47 |
| | 4.1.2 Ergebnisse der intrinsischen Evaluation der Word-Embeddings..... | 48 |
| | 4.2 Ergebnisse aus der Vorbereitung der Daten für die Hauptkomponentenanalyse | 49 |
| | 4.3 Ergebnisse Hauptkomponentenanalysen..... | 50 |
| | 4.3.1 Mittelwerte der Faktorladungen..... | 51 |
| | 4.3.2 Beschreibung der Top-10 und Bottom-10 Faktorladungen..... | 52 |
| 5 | Interpretation und Diskussion | 68 |
| | 5.1 Interpretation der Faktoren..... | 68 |
| | 5.1.1 Faktor 1 | 68 |
| | 5.1.2 Faktor 2 | 70 |
| | 5.1.3 Faktor 3 | 72 |
| | 5.1.4 Faktor 4 | 74 |
| | 5.1.5 Faktor 5 | 77 |
| | 5.1.6 Faktor 6 | 79 |
| | 5.2 Diskussion | 81 |
| | 5.3 Limitationen | 83 |
| | 5.4 Ausblick | 85 |
| | Literaturverzeichnis..... | 86 |
| | Abbildungsverzeichnis | 98 |
| | Tabellenverzeichnis..... | 99 |

| | |
|--|-----|
| Hilfsmittelverzeichnis mit Verwendungszweck..... | 100 |
| Anhang | 101 |

1 Einleitung

Dieses Kapitel dient zur Veranschaulichung der Ausgangslage der vorliegenden Arbeit. Weiter werden die Zielsetzung und Fragestellungen sowie ein knapper Überblick des Vorgehens dargestellt. Abschliessend wird die grobe Struktur der Arbeit beschrieben.

1.1 Ausgangslage

Die Wahl einer Ausbildung oder eines Berufes ist ein wichtiger Meilenstein, der starken Einfluss auf den weiteren Verlauf des Lebens eines Menschen hat (Gfrörer, Stoll, Rieger, Trautwein & Nagengast, 2022). Ein ausschlaggebender Faktor dabei ist das Berufsinteresse, welches sich im Laufe des Lebens entwickelt (Silvia, 2006). In der Schweiz beginnt der Berufswahlprozess von Jugendlichen bereits in der achten oder neunten Klasse. Während dieser Zeit setzen sich Jugendliche mit der Unterstützung durch Eltern, Schule und Berufsberatung mit ihren Fähigkeiten und Interessen auseinander (BerufsbildungPlus.ch, n.d.) Um diese Interessen besser aufdecken und präziser beschreiben zu können wurden diverse Berufswahltheorien, Interessensstrukturmodelle und Messinstrumente entwickelt (Spokane, Luchetta & Richwine, 2002). Eine der bekanntesten Theorien ist die Holland Theorie, aus welcher das RIASEC-Modell hervorging (Sodano, 2015). Holland (1997) ging dabei von den sechs Interessensdimensionen Realistic, Investigative, Artistic, Social, Enterprising und Conventional aus, mit welchen sich berufliche Interessen und Arbeitsumwelten beschreiben lassen. Trotz der etablierten Nutzung und Anerkennung des RIASEC-Modells wurde dieses im Laufe der Zeit auch kritisch hinterfragt und verschiedene Vorschläge zur Optimierung und Erweiterung des Modells wurden ausgearbeitet (Nauta, 2010; Rounds & Tracey, 1993). Trotzdem bildet das RIASEC-Modell noch immer die Basis für diverse Messinstrumente in der Berufsinteressensforschung. So entwickelten Wetzel, Hell und Pässler (2012), basierend

auf dem Modell, den Verben-Interessenstest (VIT), der berufliche Interessen mittels kontextloser Tätigkeitsverben erfasst. Diese Verben sind jeweils einer der RIASEC-Dimensionen zugeordnet und repräsentieren diese.

Gleichzeitig eröffnen die Entwicklungen der künstlichen Intelligenz (KI) und deren Anwendung in den unterschiedlichsten Bereichen neue Möglichkeiten, diese auch in den Kontext der Berufsinteressen miteinzubeziehen (Haenlein & Kaplan, 2019; Lee, 2018). Einer der zentralen Bereiche von KI ist Natural Language Processing (NLP), welche die Fähigkeit von Systemen umfasst, natürliche Sprache zu verarbeiten (Lane & Dyschel, 2025). Dabei ist das Konzept von Word-Embeddings ein NLP-Ansatz, der es Systemen ermöglicht die Bedeutung von Wörtern und ihre Beziehung zu anderen Wörtern zu verstehen. Die Wörter werden als Vektoren in einem mehrdimensionalen Raum abgebildet und jede der Dimensionen entspricht dabei einem Bedeutungsmerkmal (Hirschle, 2022). Word-Embeddings bilden nicht nur die Bedeutung und semantischen Beziehungen ab, sondern erlauben es auch latente Strukturen wie sozialen Status oder Stereotypen zu entdecken (vgl. Caliskan, Bryson & Narayanan, 2017; Kozłowski, Taddy & Evans, 2019). Diese Eigenschaft macht Word-Embeddings zu einem zunehmend relevanten Instrument in Sozialwissenschaften und weist weiter auf das Potenzial für den Einsatz in anderen wissenschaftlichen Disziplinen, wie beispielsweise der Berufsinteressenforschung hin (Arseniev-Koehler, 2024). Die vorliegende Arbeit beschäftigt sich deshalb mit der Betrachtung des RIASEC-Modells (Holland, 1997) aus der Perspektive von künstlicher Intelligenz. Insbesondere die Items des VITs bieten dabei eine vielversprechende Datenbasis, um zu untersuchen, ob und in welchem Ausmass sich die Struktur des RIASEC-Modells in den semantischen Repräsentationen eines Word-Embeddings widerspiegelt.

1.2 Zielsetzung

Wie in Kapitel 1.1 erwähnt, wurde die Berufsinteressensforschung sowie das Themenfeld der KI stark vorangetrieben. Das Ziel dieser Arbeit ist es deshalb zu untersuchen, ob und wie sich die sechs Dimensionen des RIASEC-Modells (Holland, 1997) in der Repräsentation der Tätigkeitsverben des VITs in einem Word-Embedding widerspiegelt. Im Zentrum der Arbeit stehen die Verben aus dem VIT, welche jeweils repräsentativ für die Dimensionen des RIASEC-Modells sind (Wetzel et al., 2012). Es wird überprüft ob sich die Verben über das Word-Embedding gleich oder ähnlich gruppieren, wie das durch die Fachpersonen der Berufsinteressensforschung im Rahmen des VITs gemacht wurde. Das erlaubt eine Reflexion darüber wie KI Interessensstrukturmodelle, in diesem Fall das RAISEC-Modell, erfasst und darstellt. Daraus ergeben sich die folgenden Fragestellungen:

1. *Inwiefern kann künstliche Intelligenz die menschliche Realität von Interessensstrukturmodellen insbesondere dem RIASEC-Modell widerspiegeln?*
2. *Welche neuen Dimensionen lassen sich aus dem Word-Embedding ableiten und welche Bedeutung haben diese?*

Auf Basis der Fragen und der Literatur wurden folgende Hypothesen für die Arbeit abgeleitet: *Künstliche Intelligenz spiegelt nur einen Teil der menschlichen Realität des RIASEC-Modells wider, bildet aber noch weitere eigene Dimensionen. Es lassen sich keine neuen Dimensionen aus dem Word-Embedding ableiten.*

Zur Beantwortung der Fragen und Prüfung der Hypothesen werden zuerst verschiedene deutsche Word-Embeddings evaluiert und per Ausschlussverfahren eines ausgewählt. Die Dimensionen des Word-Embeddings werden zunächst reduziert, um mit den Vektoren der Verben des VITs eine Hauptkomponentenanalyse durchzuführen. Dadurch sollen latente Strukturen sichtbar gemacht werden um diese zu analysieren.

1.3 Aufbau der Arbeit

Zunächst erfolgt in Kapitel 2 die Vorstellung des theoretischen Hintergrunds der Arbeit. Im ersten Teil der des Kapitels werden zentrale Begriffe der künstlichen Intelligenz erklärt und das Konzept von Word-Embeddings erläutert. Darauf folgt der psychologische Bezugsrahmen dieser Arbeit in Form von berufsbezogenen Interessen. Besonderer Fokus wird dabei auf das RIASEC-Modell (Holland, 1997) und den VIT (Wetzel et al., 2012) gelegt. Kapitel 3 erläutert das methodische Vorgehen zur Beantwortung der Fragestellungen. In Kapitel 4 werden die Ergebnisse präsentiert. In der Diskussion in Kapitel 5, werden die Ergebnisse zunächst interpretiert und diskutiert. Der Abschluss der Arbeit bildet das aufzeigen der Limitationen und der Ausblick auf mögliche weiterführende Forschung.

2 Theoretische Grundlagen

Dieses Kapitel bildet die theoretische Grundlage der vorliegenden Masterarbeit. Zunächst folgt eine Definition von künstlicher Intelligenz und natürlicher Sprachverarbeitung. Anschliessend liegt der Fokus auf der Erklärung des Konzeptes von Word-Embeddings, welches ein zentraler Bestandteil dieser Arbeit darstellt. Im Anschluss auf die Erklärung der technischen Konzepte und Begriffe wird auf die Thematik der interessenorientierten Berufswahl eingegangen. Der Begriffsbestimmung und der Erklärung zur Entwicklung von Interessen folgt eine Beschreibung des RIASEC-Modells (Holland, 1997) und einigen Vorschlägen zur Erweiterung dieses. Das letzte Kapitel der theoretischen Grundlage widmet sich dem Verben-Interessens-Test (VIT; Wetzel, Hell & Pässler, 2012) dessen Verbenliste ein weiterer zentraler Bestandteil dieser Arbeit bildet.

2.1 Künstliche Intelligenz und Sprachverarbeitung

Für lange Zeit war Künstliche Intelligenz (KI) eine schwer greifbare Thematik, welche nur in begrenztem Masse praktische Anwendung fand. Doch in den letzten Jahren wuchs das

Thema der künstlichen Intelligenz stark an und schaffte es, sich in der Gesellschaft zu etablieren (Haenlein & Kaplan, 2019). KI findet in immer mehr unterschiedlichen Lebensbereichen Anwendung und wirft die Frage auf, welche Wirkung sie auf Forschung und Gesellschaft haben wird (Lee, 2018). Der Begriff KI fasst eine enorme Anzahl von unterschiedlichen Themenfeldern wie maschinelles Lernen, neuronale Netzwerke und natürlicher Sprachverarbeitung zusammen (Paass & Hecker, 2024). Dieses Kapitel fokussiert sich auf die Begriffsbestimmung von KI und natürlicher Sprachverarbeitung, welche die Grundlage für das darauffolgende Konzept von Word-Embeddings bildet (Hirschle, 2022; Kaplan, 2017).

2.1.1 Künstliche Intelligenz

Erstmals erwähnt wurde der Begriff künstliche Intelligenz in einem Forschungsantrag für das Dartmouth Summer Research Project 1956 (McCarthy, Minsky, Rochester & Shannon, 1955). Für künstliche Intelligenz (KI) existiert seither noch immer keine einheitliche Definition, da bereits der grundlegende Begriff „Intelligenz“ unterschiedlich definiert wird (Mitchell, 2019; Paass & Hecker, 2024). Hinzu kommt, dass sich der Begriff „KI“ seit seiner Entstehung in den 50er Jahren immer an die fortschreitenden technischen Möglichkeiten angepasst hat (Bitkom & DFKI, 2017) Nach Russell und Norvig (2021) werden bei der Begriffsbestimmung die zwei Dimensionen menschlich vs. rational und Denken vs. Verhalten miteinbezogen. Aus der Kombination dieser Dimensionen ergeben sich vier Definitionsansätze: Menschliches Denken, menschliches Verhalten, rationales Denken und rationales Verhalten.

Menschliches Verhalten

Beispieldefinition: „Artificial Intelligence is the art of creating machines that perform functions that require intelligence when performed by people.”

(Kurzweil, 1990, S. 14)

Um dem Definitionsansatz des menschlichen Verhaltens zu entsprechen und als KI bezeichnet werden zu können, muss ein System eine Aufgabe bewältigen, welche normalerweise menschliche Intelligenz erfordert. Um beurteilen zu können ob ein System tatsächlich Intelligent ist, entwickelte Alan Turing (1950) den Turing-Test. In diesem Test kommuniziert ein System mit einem menschlichen Fragesteller. Ist für den Fragesteller unklar, ob es sich bei den Antworten um die eines Menschen oder einer Maschine handeln, so gilt der Test als bestanden. Um diesen Test zu bestehen muss ein System in der Lage sein natürliche Sprache zu verstehen, Wissen zu repräsentieren, automatisch Schlüsse zu ziehen und maschinell zu lernen (Paass & Hecker, 2024).

Menschliches Denken

Beispieldefinition: „By „think” we shall mean a performance of activities that we associate with human thinking, activities such as decision making, problem solving, learning, creating, game playing, and so on.”

(Bellman, 1978, S. 3)

Um behaupten zu können, dass KI denkt wie ein Mensch, muss erst ein Verständnis dafür vorhanden sein, wie menschliches Denken funktioniert. Dieses Verständnis kann durch Introspektion (Kuhn, 2000), psychologische Experimente (Glatzeder, Goel & Müller, 2010) oder Gehirnschans (Haynes, 2012) geschaffen werden. Ist die Theorie über menschliches Denken präzise genug, so kann sie in Form eines Computerprogrammes umgesetzt werden.

Stimmen die Input-Output-Verhaltensweisen des Systems mit dem eines Menschen überein, so kann davon ausgegangen werden, dass sich die Mechanismen ähnlich sind (Russell & Norvig, 2021).

Rationales Verhalten

Beispieldefinition: „Computational Intelligence is the study of the design of intelligent agents.” (Poole, Mackworth & Goebel, 1998, S. 1)

Der Definitionsansatz des rationalen Handelns beschreibt KI als einen rationalen Agenten. Ein Agent ist ein System, welches handelt und ein rationaler Agent handelt so, dass das bestmögliche Ergebnis, oder bei Unsicherheit das beste zu erwartende Ergebnis, erreicht wird (Wooldridge, 2000). Ein System, welches rational handelt zieht korrekte Schlüsse, geht jedoch noch darüber hinaus. Intuitive sowie reflexartige Reaktionen können rational sein, solange diese zielführend sind. Die Fähigkeiten, welche dabei eine zentrale Rolle spielen sind Wissensrepräsentation, Sprachverarbeitung, Schlussfolgerung und Lernen. Der Vorteil des rationalen Verhaltens gegenüber den anderen Ansätzen ist, dass sich Rationalität mathematisch definieren lässt. Das bedeutet auch, dass sich rationales Verhalten gezielt in die Entwicklung von Agenten einbinden lässt (Russell & Norvig, 2021).

Rationales Denken

Beispieldefinition: „Artificial intelligence is the study of mental faculties through the use of computational models.” (Charniak & McDermott, 1985, S. 6)

Mit dem Definitionsansatz des rationalen Denkens wird versucht, das korrekte Denken durch nachvollziehbare und formalisierbare Regeln zu beschreiben. Die ersten Versuche

rationales Denken in Regeln zu fassen gehen auf Aristoteles zurück. Unter wahren Voraussetzungen würden diese Regeln immer zum korrekten Schluss führen. Ein Beispiel dafür ist: „Alle Hunde sind Säugetiere. Mein Haustier ist ein Hund. Mein Haustier ist also ein Säugetier“. Solche logische Regeln dienen dazu, den Verstand abzubilden und darauf aufbauend intelligente Systeme zu entwickeln (Russell & Norvig, 2021).

Eine Definition von KI aus dem deutschsprachigen Raum von Bitkom und dem deutschen Forschungszentrum für künstliche Intelligenz (2017), lautet wie folgt: „Künstliche Intelligenz ist die Eigenschaft eines IT-Systems, „menschenähnliche“, intelligente Verhaltensweisen zu zeigen“ (S. 28). Diese Definition stammt aus einer Kombination der Ansätze „menschliches Verhalten“ (menschenähnlich) und „rationales Verhalten“ (intelligent) (Russell & Norvig, 2021). Die heutigen Anwendungs- und Forschungsbereiche der KI, in denen menschenähnliches und intelligentes Verhalten gezeigt werden soll, sind vielfältig und breiten sich weiter in unterschiedliche Domänen aus (Goodfellow, Courville & Bengio, 2016). Diese Bereiche lassen sich in verschiedene Teilbereiche gliedern, welche sich mit allgemeinen und doch komplexen Problemen befassen oder unterschiedliche Fertigkeiten und Hilfsmittel erfordern. Zu den bekannteren Bereichen von KI zählen Computer Vision, Robotik, Spracherkennung und die Verarbeitung von natürlicher Sprache (Kaplan, 2017).

2.1.2 Natural Language Processing

Die Verarbeitung von natürlicher Sprache, im englischen auch bekannt als Natural Language Processing (NLP), befasst sich mit der Entwicklung von Systemen, die menschliche Sprache verarbeiten und verstehen können (Sowmya, Majumder, Gupta & Surana, 2020). Zwischen natürlichen Sprachen, wie Deutsch oder Englisch und Programmiersprachen wie Python oder Java existieren gewisse Parallelen. Um verständliche Aussagen formulieren zu können, folgen beide grammatikalischen Regeln, beide bestehen aus sogenannten „Token“, also Einheiten, die eine Bedeutung tragen. Trotz dieser Parallelen sind

die Sprachen grundlegend verschieden. Programmiersprache ist künstlich konstruiert, und dient dazu exakt definierte Abläufe auf Maschinenebene auszuführen. Sie ist immer eindeutig und ein Input kann immer nur auf eine bestimmte Weise interpretiert werden. Natürliche Sprachen hingegen sind evolutionär entstanden und dienen zur Verständigung zwischen Menschen, die kognitive Fähigkeiten besitzen. Sie kann mehrdeutig sein und Wörter können je nach Kontext unterschiedliche Bedeutung haben. Dieser Unterschied stellt eine Herausforderung dar, wenn natürliche Sprache maschinell verarbeitet werden soll. Dort setzt das Forschungsfeld Natural Language Processing an (Lane & Dyshel, 2025). Die Fähigkeit natürliche Sprache maschinell zu verarbeiten, ermöglicht es Computern sprachbezogene Aufgaben zu bewältigen. Ein typisches Beispiel einer solchen Aufgabe sind sogenannte Conversational Agents wie *ChatGPT*. Systeme wie diese beinhalten unter anderem Spracherkennung, Sprachverständnis, Sprachgenerierung und Sprachsynthese (Jurafsky & Martin, 2014). Um es Maschinen zu ermöglichen natürliche Sprache zu verarbeiten, muss diese zuerst in eine für sie lesbare Form umgewandelt werden – meist in Form von Zahlen. Eine Methode dazu ist der Bag-of-Words-Ansatz. Dieser codiert Wörter komplett unabhängig voneinander. Ein Problem dieses Ansatzes ist, dass dieser inhaltlich unpräzise wird, da die Worte in keinerlei Relation zueinander stehen. Semantische Beziehungen und grammatikalische Übereinstimmungen werden somit nicht repräsentiert. Um dieses Problem zu lösen wird das Verfahren der Worteinbettung (Word-Embedding) genutzt (Hirsche, 2022).

2.2 Word-Embeddings

Aufbauend auf den Grundlagen der künstlichen Intelligenz und der natürlichen Sprachverarbeitung wird im folgenden Kapitel das Konzept der Word-Embeddings dargestellt. Word-Embeddings bilden eine zentrale Methode, um sprachliche Bedeutungen in numerischer Form abzubilden (Sowmya et al., 2020). Das Kapitel erläutert zunächst die Prinzipien von Word-Embeddings und stellt anschliessend die Word-Embedding Modelle

Word2Vec (Mikolov, Sutskever, Chen, Corrado & Dean, 2013), GloVe (Pennington, Socher & Manning, 2014) und fastText (Joulin, Grave, Bojanowski & Mikolov, 2017) vor.

2.2.1 Grundlagen von Word-Embeddings

Das Grundprinzip von Word-Embeddings besteht darin, die Bedeutung von Wörtern und die Beziehung, welche sie zu anderen Wörtern haben, mithilfe von Vektoren darzustellen. Jedes Wort wird in einem mehrdimensionalen Raum als Punkt abgebildet. Jede dieser Dimensionen spiegelt eine Art Bedeutungsmerkmal wider, wie beispielsweise semantische Nähe oder grammatikalische Eigenschaften. Aus der Kombination der Werte dieser Dimensionen ergibt sich die Position eines Wortes in diesem Raum. Somit liegt zwischen Wörtern, welche eine ähnliche Bedeutung haben eine geringe Distanz, während unterschiedliche Wörter weit voneinander entfernt sind (Hirschle, 2022). Ein Beispiel mit den Wörtern „Hund“, „Katze“ und „Auto“ illustriert die Funktionsweise von Word-Embeddings. Während „Hund“ und „Katze“ semantische Relationen haben, da beides Haustiere sind, würden diese Wörter im mehrdimensionalen Raum nahe beieinander sein. Das Wort „Auto“ hingegen steht in keiner semantischen Relation zu den Wörtern „Hund“ und „Katze“ und würde demnach weit von ihnen entfernt positioniert sein. Die Abbildung 1 zeigt dieses Beispiel stark vereinfacht in einem zweidimensionalen Raum.

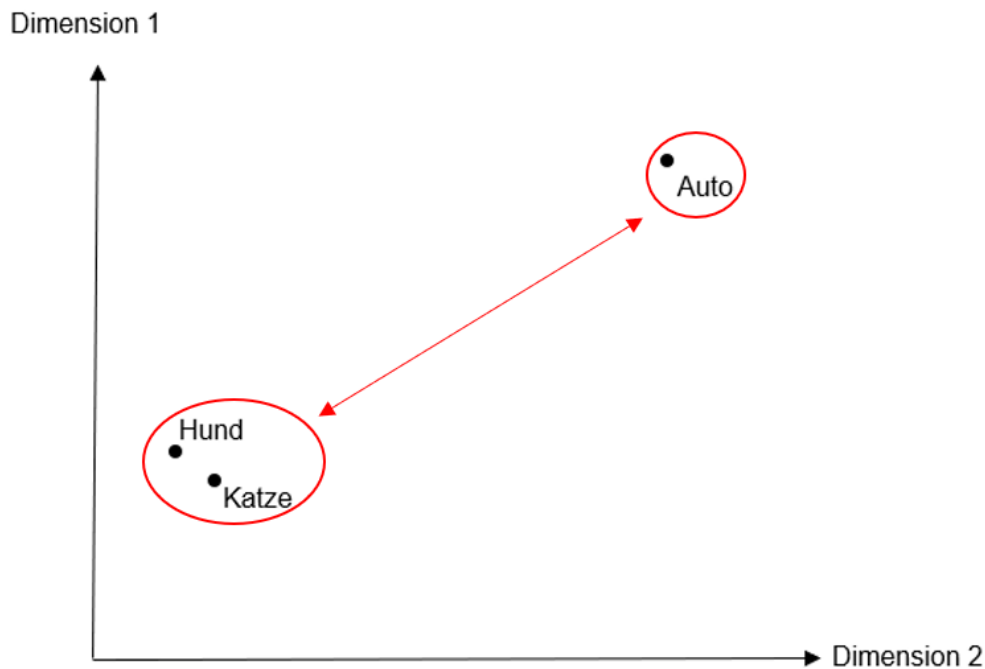


Abbildung 1. Illustration zur Funktionsweise von Word-Embeddings in einem zweidimensionalen Raum (eigene Darstellung)

Die Bedeutungsdimensionen von Word-Embeddings werden nicht manuell definiert, sondern werden auf Basis der Verwendung der Wörter aus einem Trainingsdatensatz berechnet. Bei solchen Trainingsdatensätzen handelt es sich um Textdaten, welche beispielsweise aus Nachrichtenartikeln, Wikipedia-Einträgen, Social-Media-Beiträgen oder sonstigen Dokumenten bestehen (Arseniev-Koehler, 2024). Die Bedeutung der Worte und somit auch die Bedeutungsdimensionen basieren auf der distributionalen Hypothese, aus der Linguistik. Diese besagt, dass Worte, die in ähnlichen Kontexten auftauchen, auch eine ähnliche Bedeutung haben (Firth, 1968, zitiert nach Sowmya et al., 2020, S. 92). Die Anzahl der Dimensionen eines Word-Embeddings werden durch den Algorithmus oder Anwendender*in zuvor festgelegt. Für gewöhnlich bewegt sie sich im Bereich von mehreren hundert Dimensionen (Rong, 2014). Dabei ist die Bedeutung der einzelnen Dimensionen nicht ersichtlich und für Menschen nicht direkt interpretierbar. Während des Trainingsprozesses werden die Dimensionen durch den Algorithmus festgelegt. Dies geschieht, indem die Wörter des Trainingsdatensatzes räumlich angeordnet werden (Arseniev-

Koehler, 2024). Es existieren unterschiedliche Trainingsalgorithmen für Word-Embeddings, wie fastText (Joulin et al., 2017), GloVe (Pennington et al., 2014) und Word2Vec (Mikolov, Sutskever, et al., 2013), auf diese wird in Kapitel 2.2.2 näher eingegangen. Wie bereits erwähnt drückt der Abstand zwischen Wörtern im Raum die Ähnlichkeit dieser Wörter aus. Diese Ähnlichkeit wird meist durch die Kosinus-Ähnlichkeit berechnet (Arseniev-Koehler, 2024). Ein gut trainiertes Word-Embedding ist in der Lage, semantische Informationen so zu erfassen, dass es mit dem menschlichen Wortverständnis vergleichbar ist (Mandera, Keuleers & Brysbaert, 2017). Beispielsweise stimmen die durch Word-Embeddings berechneten Kosinus-Ähnlichkeiten zwischen Wortvektoren deutlich mit den durch Menschen eingeschätzten Bedeutungsähnlichkeiten von Wörtern überein (Caliskan et al., 2017). Nicht nur die Position der einzelnen Wörter im Vektorraum, sondern auch die Richtung zwischen ihnen können relevante Informationen für deren Bedeutung enthalten. Ein bekanntes Beispiel dazu ist die semantische Beziehung der Wörter „König“ und „Königin“ in Bezug auf die Wörter „Mann“ und „Frau“. Die Differenz in den Wortvektoren von „König“ und „Königin“ stimmt oft mit der Differenz aus den Wortvektoren von „Mann“ und „Frau“ überein (Mikolov, Chen, Corrado & Dean, 2013). Das bedeutet, dass die Differenz zwischen diesen Wortvektoren ein Vektor ist, welcher wie eine Linie im Raum zu interpretieren ist. Diese Linie gibt eine Richtung an, welche sich als eine Art Bedeutungsachse verstehen lässt. Bei dieser Achse zeigt ein Ende in die Richtung von weiblich konnotierten Begriffen und das andere Ende in die Richtung von männlich konnotierten Begriffen (Arseniev-Koehler, 2024). Diese Bedeutungsachsen beschränken sich nicht nur auf das Geschlecht. Es lassen sich dadurch auch andere latente Konzepte erfassen, wie beispielsweise sozialer Status (Kozlowski et al., 2019) oder Stereotypen (Caliskan et al., 2017). Diese Eigenschaft von Word-Embeddings machen sie besonders interessant für den Einsatz in Sozialen Wissenschaften, aber auch in anderen wissenschaftlichen Disziplinen (Arseniev-Koehler, 2024).

2.2.2 Unterschiedliche Word-Embedding Modelle

Die Vektorisierung von Wörtern ist ein zentrales Verfahren bei der Erstellung von Word-Embeddings. In früheren Verfahren zur Vektorisierung wurden die Modelle jeweils für eine spezifische NLP-Aufgabe trainiert. Daran problematisch war jedoch die hohe Rechenintensität. Es mussten mindestens so viele Parameter für jedes Wort trainiert werden, wie es Dimensionen im Modell gab. Ein weiteres Problem waren die Trainingsdaten. Diese waren oft unvollständig oder qualitativ unzureichend. Das führte dazu, dass einige Wörter nur teilweise oder sogar gar nicht trainiert wurden. Aus diesem Grund wurden neue Verfahren entwickelt, welche die Bedeutung von Wörtern anhand ihrer Verwendung in Kontexten repräsentieren, um einen semantischen Raum zu erlernen. Das Ziel dieser Verfahren ist es, ein allgemeines Sprachmodell zu schaffen (Hirschle, 2022). Einige der bekannteren neuen Modelle sind Word2Vec (Mikolov, Sutskever, et al., 2013), GloVe (Pennington et al., 2014) und fastText (Joulin et al., 2017).

2.2.2.1 Word2Vec

Das von Mikolov, Sutskever et al. (2013) entwickelte Modell Word2Vec basiert auf einem einfachen neuronalen Netzwerk. Dazu wurde das Skip-Gram-Verfahren oder continuous-bag-of-words-Verfahren (CBOW) genutzt. Skip-Gram versucht anhand eines vorgegebenen Wortes die Wörter aus der Umgebung vorherzusagen. Während CBOW das umgekehrte Prinzip verfolgt, das Zielwort anhand der Wörter aus der Umgebung vorherzusagen (Mikolov, Chen, et al., 2013). Diese Verfahren beruhen auf der Annahme, dass der natürliche Kontext, in dem Wörter im Sprachgebrauch eingebettet sind, Rückschlüsse auf ihre Bedeutung zulassen (Hirschle, 2022). Der Unterschied zu früheren neuronalen Netzwerken ist, dass keine dichten Matrixmultiplikationen verwendet werden, was das Training deutlich effizienter macht (Mikolov, Chen et al., 2013). Zum Training eines Word2Vec-Modells wird ein Textkorpus mit Sätzen verwendet, die diesen natürlichen

Kontext wiedergeben. Während des Trainings lernt das Modell die Beziehung zwischen einem Wort und den Nachbarwörtern (Hirschle, 2022). Das Modell soll durch das angepasste Verfahren Wortrepräsentationen erzeugen, welche die syntaktische, als auch die semantische Beziehung zwischen Wörtern abbildet. Viele dieser Beziehungen zeigen sich überraschenderweise als lineare Zusammenhänge im Vektorraum. Wird beispielsweise der Vektor „Spanien“ vom Vektor „Madrid“ subtrahiert und anschliessend der Vektor „Frankreich“ addiert so ist das Ergebnis näher am Vektor „Paris“ als an anderen Wortvektoren. Auch einfache Additionen führen zu sinnvollen Ergebnissen. Die Addition der Wortvektoren von „Deutschland“ und „Hauptstadt“ ergeben einen Vektor in der Nähe des Wortvektors „Berlin“. Diese linearen Zusammenhänge erlauben es dem Modell analoge Schlussfolgerungen zu ziehen (Mikolov, Sutskever, et al., 2013).

Schlussendlich wurde das Modell erweitert, um auch Phrasen wie „Chicago Bears“ abbilden zu können. „Chicago Bears“ ist der Name eines Football-Teams und somit keine natürliche Kombination aus den Worten „Chicago“ und „Bears“. Die Erweiterung erfolgte, in dem eine grosse Menge an Phrasen identifiziert wurde, um dann als eigene Einheiten in das Training miteinbezogen zu werden (Mikolov, Sutskever, et al., 2013).

2.2.2.2 GloVe

Pennington et al. (2014) entwickelten das GloVe-Modell (Global Vectors for Word Representation), welches einen anderen Ansatz als Word2Vec verfolgt. Statt einzelne Worte anhand ihres lokalen Kontextes vorherzusagen, basiert GloVe auf der Auswertung globaler Kookurrenzstatistiken. Das bedeutet, dass analysiert wird, wie häufig Wörter zusammen in einem festgelegten Kontext in den Trainingsdaten vorkommen. Die Idee ist, dass diese Häufigkeit Informationen über die Bedeutung der Wörter beinhaltet (Toshevskaja, Stojanovskaja & Kalajdjieski, 2020).

Dabei zentral ist, dass nicht nur die absoluten Häufigkeiten analysiert werden, sondern auch das Verhältnis wie oft ein Wort im Gegensatz zu einem anderen Wort in Kombination mit einem dritten Wort vorkommt (Pennington et al., 2014). Ein anschauliches Beispiel zu den Kookkurrenzverhältnissen lässt sich mit den Wörtern „Hund“, „Tiger“ und „Haustier“ zeigen. Die Kombination der Wörter „Hund“ und „Haustier“ kommt in einem fiktiven Trainingsdatensatz 200-mal vor. Diese Häufigkeit deutet bereits auf eine semantische Beziehung zwischen den Wörtern „Hund“ und „Haustier“ hin. Die Kombination der Wörter „Tiger“ und „Haustier“ kommt hingegen 100-mal vor. Auch dies könnte ein Hinweis auf eine semantische Relation der beiden Wörter sein. Um jedoch zu erkennen, welche Wörter semantisch näher sind untersucht GloVe zusätzlich das Verhältnis der Kookkurrenzen. Die Kombination der Wörter „Hund“ und „Haustier“ kommt doppelt so oft vor, wie die Kombination der Wörter „Tiger“ und „Haustier“. Daraus lässt sich schliessen, dass das Wort „Hund“ eine viel engere Beziehung zum Wort „Haustier“ hat, als das Wort „Tiger“.

Die Wortvektoren werden so berechnet, dass der mathematische Zusammenhang möglichst gut zu der Häufigkeit passt, mit welcher die Wörter gemeinsam in den Trainingsdaten vorkommen. Zusätzlich wird ein Korrekturwert für jedes Wort miteinbezogen. Dieser soll Unterschiede in der Häufigkeit, mit welcher Wörter im Trainingsdatensatz vorkommen, ausgleichen. Beispielsweise kommen Wörter wie „und“ oder „der“ sehr oft im Trainingsdatensatz vor, liefern aber wenig inhaltliche Information. Eine zusätzliche Gewichtungsfunktion sorgt dafür, dass sehr häufige oder extrem seltene Wortpaare die Ergebnisse nicht verzerren (Pennington et al., 2014).

2.2.2.3 fastText

Das fastText-Modell von Joulin et al. (2016) ist eine Weiterentwicklung des Word2Vec-Algorithmus und basiert ebenfalls auf dem Skip-Gram oder CBOW-Verfahren (Hirschle, 2022). Anders als bei Word2Vec werden aber nicht ganze Wörter als

Recheneinheit verwendet, sondern sie werden in kleinere Bestandteile zerlegt. Diese Bestandteile werden Subwörter oder auch n-grams genannt und sind Buchstabenfolgen, welche aus drei bis sechs Buchstaben bestehen. Die Summe der Vektoren dieser n-grams repräsentieren dann das gesamte Wort. Das ermöglicht dem Modell auch aus den Bestandteilen eines Wortes und nicht nur aus vollständigen Wörtern ihre Bedeutung abzuleiten. Zum Beispiel wird das Wort „where“ in die Einzelteile <„wh“, „whe“, „her“, „ere“, „re“> aufgeteilt. Um den Anfang und das Ende eines Wortes zu markieren, werden spitze Klammern verwendet (Bojanowski, Grave, Joulin & Mikolov, 2017). Ein Vorteil dieses Modelles zeigt sich insbesondere darin, dass es in der Lage ist, für unterschiedliche Wortformen und für seltene und zusammengesetzte Wörter, sinnvolle Vektoren zu berechnen. Das macht fastText besonders interessant für deutsche Word-Embeddings, da in der deutschen Sprache viele zusammengesetzte Wörter existieren (Hirschle, 2022).

2.3 Berufsbezogene Interessen

Nach der Betrachtung der technologischen Grundlagen und Theorie wird im folgenden Kapitel der psychologische Bezugsrahmen dieser Arbeit dargestellt. Es folgt zuerst eine Einführung in den Begriff von Interesse, welches das übergeordnete Forschungsthema der Berufsinteressen bildet (Silvia, 2006). Das RIASEC-Modell von Holland (1997) gilt als das dominanteste Modell für berufsbezogene Interessen. Auf dieses wird in diesem Kapitel ausführlich eingegangen, da es die theoretische Grundlage des VITs darstellt. Der Beschreibung des Modells folgt ein Kapitel mit Erweiterungsvorschlägen aus der Berufsinteressensforschung. Einen besonderen Fokus dieses Kapitels widmet sich dem VIT, zumal dessen Verbenliste die Grundlage der analysierten Daten bildet (Wetzel et al., 2012).

2.3.1 Interessen

In der Literatur existiert trotz reichlicher Forschung keine einheitliche Definition des Begriffs Interesse (vgl. Hansen, 2013; Krapp & Prenzel, 2011; Mount et al., 2005; Silvia, 2006). Laut Silvia (2006) lassen sich die Theorien zu Interesse in zwei Bereiche aufteilen. Einerseits lässt sich Interesse als eine momentane Motivation, Neugier oder emotionale Erfahrung verstehen. Andererseits ist Interesse ein Bestandteil der Persönlichkeit, durch die sich Menschen unterscheiden, beispielsweise durch individuelle Ziele, Neigungen und Hobbys. Diese Unterscheidung der Bereiche stimmt auch mit der Aufteilung von Interesse in situationales und individuelles Interesse nach Hidi (1990) überein. Krapp und Prenzel (2011) definieren Interesse aufgrund des Fokus auf spezifische Inhalte oder Gegenstände als die Tendenz, mit welcher Personen sich bestimmten Themen, Objekten und Kontexten zu- oder abwenden. Damit im Zusammenhang steht auch die Personen-Gegenstands-Theorie, welche Interesse als eine ungefähr überdauernde und spezifische Beziehung zwischen einem Objekt und einer Person versteht. Dabei bezieht sich der Gegenstand nicht nur auf ein physisches Objekt, sondern kann auch eine Idee oder Themenbereich darstellen. Auch diese Theorie lässt sich von den beiden Perspektiven des situationalen und individuellen Interesses betrachten (Krapp, 2002). Holland (1997) versteht unter Interesse den Ausdruck der eigenen Persönlichkeit durch Vorlieben, Arbeit und Gestaltung der Freizeit. Demgegenüber interpretieren Hogan und Blake (1999) Interesse eher als eine Spiegelung der eigenen Identität als einen Ausdruck der Persönlichkeit. Dabei beschreibt sich die Identität dadurch wie eine Person über sich selbst denkt und was ihr wichtig ist wie Ziele, Werte und Bestrebungen. Trotz diverser unterschiedlicher Erklärungsversuche gibt es gemäss Renninger und Hidi (2019) in der Interessensforschung einen weitverbreiteten Konsens zu folgenden Charakteristiken von Interesse: Interesse besitzt eine kognitive und affektive Komponente, entsteht auf einer biologischen Basis, existiert innerhalb der Interaktion zwischen einer Person

und ihrer Umwelt, kann sowohl bewusst als auch unbewusst sein und bezieht sich immer auf spezifische Inhalte.

Personen werden jedoch nicht mit ihren Interessen geboren, sondern entwickeln diese im Verlaufe des Lebens (Silvia, 2006). Hidi & Renninger (2006) beschreiben die Entwicklung von Interessen in vier Phasen: Die erste Phase beschreibt ein ausgelöstes situationales Interesse. Dabei wird dieses typischerweise durch einen externen Reiz ausgelöst. Falls dieses situationale Interesse aufrechterhalten wird, entwickelt sich daraus ein bestehendes situationales Interesse, was die zweite Phase des Modells bildet. Die Aufrechterhaltung des situationalen Interessens erfolgt dabei meist durch äussere Einflüsse. In der dritten Phase entsteht aus dem bestehenden situationalen Interesse ein schwach entwickeltes, individuelles Interesse. Dieses zeichnet sich insbesondere durch eine wiederholte freiwillige Auseinandersetzung mit Inhalten aus. Bei entsprechender Förderung beginnt die vierte Phase und das individuelle Interesse kann sich komplett entwickeln und festigen. Ein voll entwickeltes Interesse charakterisiert sich durch eine stabile, wiederkehrende freiwillige Auseinandersetzung, einem tiefem Wissensstand und einer persönlichen Bedeutung. Die vier Phasen des Modells sind dabei immer aufeinanderfolgend und klar voneinander getrennt. Wie lange eine Phase dauert oder wie ausgeprägt diese ist, unterscheidet sich dabei je nach Person (Hidi & Renninger, 2006).

Die individuellen Interessen stellen ein entscheidendes Auswahlkriterium für Personen dar, um einen passenden, erfüllenden und zufriedenstellenden Beruf zu finden. Die Forschung der berufsbezogenen Interessen gilt dabei als das vermutlich am stärksten beforschte Gebiet der Interessenspsychologie (Silvia, 2006). Aus dieser Forschung gehen diverse Modelle und Theorien zu Berufsinteressen und deren Erfassung hervor (Sodano, 2015).

2.3.2 *Hollands Theorie (RIASEC)*

Eine der wohl bekanntesten Theorien zu Berufsinteressen ist Hollands Typentheorie aus dem Jahr 1985, welche davon ausgeht, dass die Wahl eines Berufs ein Ausdruck der eigenen Persönlichkeit sei (Sharf, 2013). Dabei fokussiert sich die Theorie hauptsächlich auf Interessen, inkludiert aber auch die mit dem Beruf verbundenen Fähigkeiten, Fertigkeiten, Wirksamkeiten und Werte (Sodano, 2015). Eine zentrale Rolle dabei spielt die Passung zwischen Persönlichkeit und Arbeitsumwelt (Person-Environment fit), welche auch Kongruenz genannt wird. Es wird angenommen, dass die Zufriedenheit mit dem Beruf, die Stabilität der Zufriedenheit und der berufliche Erfolg mit der Passung zwischen Persönlichkeit und Arbeitsumwelt zusammenhängen (Nauta, 2010). Personen sind also motiviert, nach einer Arbeitsumwelt zu suchen, welche es ihnen erlaubt, ihre Fertigkeiten, Problemlösestrategien und Fähigkeiten einzusetzen, ihre Einstellungen und Werte auszudrücken und die bevorzugten Rollen anzunehmen (Sodano, 2015).

Holland (1997) geht davon aus, dass sich sowohl die beruflichen Interessen als auch die Arbeitsumwelt mittels sechs Dimensionen beschreiben lassen. Diese Dimensionen werden meist als Hexagon dargestellt (siehe Abbildung 2) und umfassen Realistic (R), Investigative (I), Artistic (A), Social (S), Enterprising (E) und Conventional (C). Das Hexagon ist so angeordnet, dass der Grad der Ähnlichkeit zwischen Dimensionen oder auch Konsistenz genannt, durch die räumliche Distanz dargestellt wird. Je näher Dimensionen beieinander liegen, desto ähnlicher sind sich diese und je weiter entfernt sie voneinander sind, desto stärker unterscheiden sie sich. Das bedeutet beispielsweise, dass sich die Dimensionen Realistic und Investigative ähnlicher sind, als die Dimensionen Realistic und Artistic. Die grösste Distanz und damit der grösste Unterschied besteht folglich zwischen Dimensionen, die sich auf dem Hexagon gegenüberliegen (Holland, 1997). Aus den Anfangsbuchstaben dieser sechs Dimensionen bildet sich das Acronym RIASEC, weshalb Holland's Theorie auch unter dem Namen RIASEC-Modell bekannt ist (Nauta, 2010; Sodano, 2015).

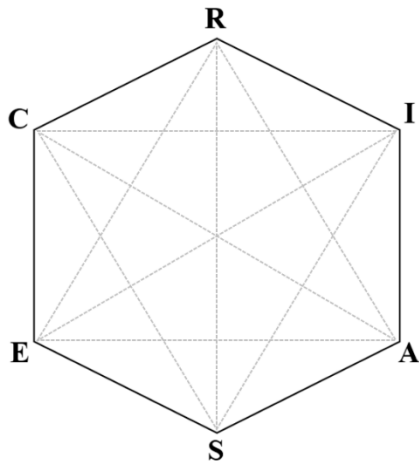


Abbildung 2. Hexagon des RIASEC-Modells nach Holland (1997) (eigene Darstellung)

Jede dieser Dimensionen steht für verschiedene Tätigkeitspräferenzen und Abneigungen. Die Realistic-Interessen beschreiben klar definierte und zielgerichtete Tätigkeiten mit Gegenständen oder Werkzeugen. Dies kann im Rahmen von technischer, mechanischer, elektrischer, manueller oder landwirtschaftlicher Arbeit gemeint sein. Personen mit Realistic-Interessen werden als praktisch und bodenständig bezeichnet und zeigen häufig eine Abneigung gegenüber erzieherischen oder sozialen Tätigkeiten (Eder & Bergmann, 2015; Holland, 1997; Sodano, 2015).

Die Investigative-Interessen beziehen sich auf Tätigkeiten, bei denen sich systematisch, symbolisch und beobachtend mit physikalischen, biologischen oder kulturellen Phänomenen auseinandergesetzt wird. Personen mit diesen Interessen werden als neugierig und wissenschaftlich beschrieben. Tätigkeiten, welche sozial, beeinflussend oder repetitiv sind, werden eher abgelehnt (Eder & Bergmann, 2015; Holland, 1997; Sodano, 2015).

Die Artistic-Interessen stehen für freie und unsystematische Tätigkeiten, bei denen Materialien, Sprache oder Darstellung zu künstlerischen Zwecken eingesetzt werden. Personen mit Artistic-Interessen werden als kreativ und unkonventionell bezeichnet. Systematische, geordnete sowie klar definierte Tätigkeiten missfallen ihnen (Eder & Bergmann, 2015; Holland, 1997; Sodano, 2015).

Mit Social-Interessen lassen sich Tätigkeiten beschreiben, bei denen andere beeinflusst werden. Dies kann beispielsweise die Information, Sensibilisierung, Ausbildung, Entwicklung oder Pflege von anderen Personen sein. Charakterisiert werden Personen mit Social-Interessen durch ihre Hilfsbereitschaft und Geselligkeit. Tätigkeiten, die systematisch, geordnet oder klar definiert sind und den Umgang mit Maschinen, Werkzeugen oder Materialien erfordern, sprechen Personen mit Social-Interessen weniger bis gar nicht an (Eder & Bergmann, 2015; Holland, 1997; Sodano, 2015).

Die Enterprising-Interessen umfassen Tätigkeiten, bei denen durch die Beeinflussung Anderer geschäftliche Ziele und wirtschaftlicher Gewinn erreicht werden. Personen mit Enterprising-Interessen werden als überzeugend und ambitiös beschrieben. Sie zeigen häufig eine Abneigung gegenüber systematischen, symbolischen und beobachtenden Tätigkeiten (Eder & Bergmann, 2015; Holland, 1997; Sodano, 2015).

Zu Conventional-Interessen gehören Tätigkeiten, bei welchen klar definiert, geordnet und systematisch mit Daten umgegangen wird. Personen, mit Conventional-Interessen charakterisieren sich durch Sorgfalt und Konformität. Tätigkeiten die unklar, frei oder unsystematisch sind, werden eher abgelehnt (Eder & Bergmann, 2015; Holland, 1997; Sodano, 2015).

Nicht nur die Interessentypen sondern auch die Arbeitsumwelten lassen sich mit diesen Dimensionen beschreiben. Wie eingangs erwähnt, streben Personen danach, eine Passung zwischen ihrer Persönlichkeit und der Arbeitsumwelt zu erreichen. Personen mit einer Ausprägung auf einer Dimension der RIASEC-Modells fühlen sich demnach von einer Arbeitsumwelt angezogen, welche die gleiche oder mindestens eine ähnliche Ausprägung aufweist (Sodano, 2015).

Eine Person kann Ausprägungen auf einer, mehreren oder sogar allen Dimensionen aufweisen (Spokane et al., 2002). Auf der Basis von Assessments werden diese Ausprägungen ermittelt und mit einem sogenannten Holland-Code dargestellt. Dieser Code

besteht aus den Anfangsbuchstaben der Dimensionen in absteigender Reihenfolge der Ausprägung. Ebenso werden verschiedene Arbeitsumwelten anhand dieser Codes beschrieben (Nauta, 2010). Typischerweise setzt sich für Beurteilung und Intervention der Holland-Code aus den drei Dimensionen mit der höchsten Ausprägung zusammen. Beispielsweise wäre der Holland-Code für einen Programmierer IRC. Diese Person hat demnach die höchste Ausprägung in der Dimension Investigative, gefolgt von Realistic und Conventional (Spokane et al., 2002). Es existieren Online-Register, in welchen anhand der Holland-Codes nach passenden Berufen gesucht werden kann. Eines der umfangreichsten dieser Register ist O*NET OnLine, welches für vom US-Arbeitsministerium angeboten wird (O*NET OnLine, 2025).

Bis heute bildet das RIASEC-Modell die Basis für zahlreiche berufsbezogene Interessenstests (Spokane et al., 2002). Dazu gehören unter anderem die Self-Directed Search (SDS; Holland, Fritzsche & Powell, 1994, zitiert nach Nauta, 2010, S. 12), das Vocational Preference Inventory (VPI; Holland, 1985, zitiert nach Spokane et al., 2002, S. 391) und das Position Classification Inventory (PCI; Gottfredson & Holland, 1991, zitiert nach Nauta, 2010, S. 12). Auch im deutschsprachigen Raum gibt es verschiedene berufsbezogene Interessenstest, die auf der Theorie von Holland basieren, wie beispielsweise Explorix (Jörin, Stoll, Bergmann & Eder, 2003), allgemeiner Interessen-Strukturtest (AIST-R; Bergmann & Eder, 2005) oder Verb-Interessentest (VIT; Wetzel, Hell & Pässler, 2012). Trotz der weitverbreiteten Anwendung existiert auch Kritik am RIASEC-Modell (Nauta, 2010). Die sechs Dimensionen des RIASEC-Modells würden nicht ausreichen, um die Komplexität und Diversität der heutigen Berufswelt zu repräsentieren (Armstrong, Day, McVay & Rounds, 2008). Das RIASEC-Modell wurde bereits in den 90er Jahren entwickelt, seither hat sich die Art wie Menschen arbeiten verändert und diverse neue Berufe kamen auf, während ältere Berufe verschwanden (Hirschi, 2018). Diverse Studien schlagen daher ein Modell mit mehr als sechs Dimensionen vor (vgl. Su, Tay, Liao, Zhang & Rounds, 2019; Wang et al., 2024).

Ein weiterer Kritikpunkt ist die mangelnde Differenzierung und die inhaltliche Weite der Dimensionen. Stark unterschiedliche Berufe werden teilweise unter den gleichen Dimensionen zusammengefasst (Armstrong, Smith, Donnay & Rounds, 2004). So finden sich im Berufsregister des AIST-R (Bergmann & Eder, 2005) unter dem Dimensionscode SRE die Berufe Servierkraft, Segellehrer*in und Fusspfleger*in.

Entgegen widersprechenden Befunden und Kritik zu unterschiedlichen Aspekten des RIASEC-Modells, ist es allgemein anerkannt und konnte sich über die Jahre sowohl in der Forschung als auch in der Praxis etablieren (Juntunen, Motl & Rozzi, 2019). Das Modell erhielt im Feld der Berufsinteressensforschung grosse Aufmerksamkeit und wurde intensiv beforscht (Hartmann, Heine & Ertl, 2021). Dadurch entstanden diverse Vorschläge zur Optimierung und Erweiterung des Modells (Rounds & Tracey, 1993).

2.3.3 Erweiterungen des RIASEC-Modells

Holland ging beim RIASEC-Modell von einer hexagonalen oder kreisförmigen Anordnung aus, bei der die Distanz zwischen den Dimensionen ihre Ähnlichkeit zueinander ausdrückt. Je weiter die Entfernung zwischen den Dimensionen, desto weniger ähnlich sind sich diese. Dabei bleibt die Distanz zwischen den Dimensionen konstant. Zwischen benachbarten Dimensionen ist eine kleine, zur übernächsten Dimension eine mittlere und zur gegenüberliegenden Dimension eine grosse Distanz (Rounds & Tracey, 1993). Die Gleichverteilung dieser Abstände erfüllt die mathematischen Anforderungen für ein Circumplex-Modell, das die Beziehung zwischen Variablen durch eine zirkuläre Anordnung abbildet. (Guttman, 1954, zitiert nach Rounds, Tracey & Hubert, 1992, S. 241). Diese kreisförmige Struktur des RIASEC-Modell erlaubt, das Modell durch zugrunde liegende Dimensionen zu ergänzen, wie es Prediger (1982) und Hogan (1983, zitiert nach Sodano, 2015, S. 284) vorgeschlagen haben. Predigers (1982) Vorschlag beinhaltet dabei zwei bipolare Dimensionen mit den Achsen Dinge-Menschen und Daten-Ideen (siehe Abbildung

3). Die Dinge-Menschen-Achse unterscheidet zwischen den dingeorientierten Dimensionen Realistic, Investigative und Conventional und den menschenorientierten Dimensionen Social, Enterprising und Artistic. Die Dimension Realistic und Social fallen dabei auf die Pole dieser Achse und weisen den stärksten inhaltlichen Zusammenhang mit dieser auf. Die Daten-Ideen-Achse unterscheidet hingegen zwischen den datenorientierten Dimensionen Conventional und Enterprising und den ideenorientierten Dimensionen Investigative und Artistic. Die vier Dimensionen hängen dabei gleich stark mit der Achse zusammen, während die Dimensionen Realistic und Social keinen Zusammenhang mit dieser aufweisen (Prediger, 1982).

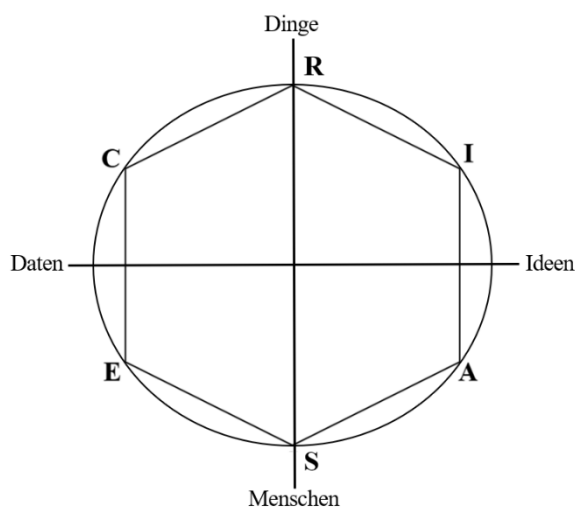


Abbildung 3. Predigers (1982) zugrundeliegenden Dimensionen (eigene Darstellung)

Hogan (1983, zitiert nach Rounds & Tracey, 1993, S. 876) schlägt eine 30 Grad Rotation von Predigers (1982) Achsen vor und benennt diese Sociability- und Conformity-Achse (siehe Abbildung 4). Dadurch wird das RIASEC-Modell mit den beiden Big Five Persönlichkeitsmerkmalen Extraversion und Offenheit in Verbindung gebracht (Rounds & Tracey, 1993). Die Sociability-Achse unterscheidet dabei zwischen den Dimensionen Enterprising und Social am oberen und den Dimensionen Realistic und Investigative am unteren Pol. Die Conformity-Achse differenziert hingegen zwischen der Dimension Conventional am oberen und der Dimension Artistic am unteren Pol (Hogan, 1983, zitiert nach Sodano, 2015, S. 284).

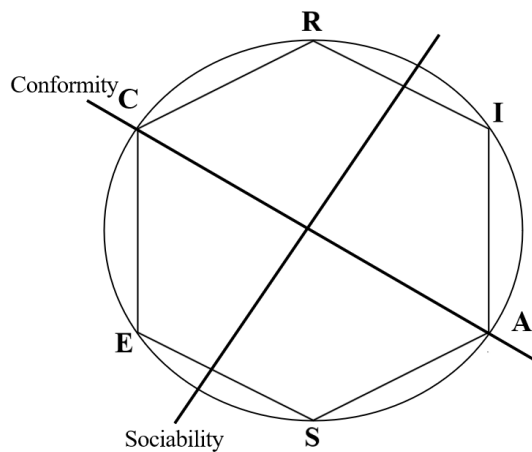


Abbildung 4. Hogans (1983, zitiert nach Sodano, 2015, S. 284) zugrundeliegenden Dimensionen (eigene Darstellung)

Hogans Vorschlag wurde in der Interessensforschung wenig Beachtung geschenkt (Rounds et al., 1992). Die zugrundeliegenden Dimensionen von Prediger (1982) gehört jedoch zu einer der bekanntesten Erweiterungen des RIASEC-Modells (Sodano, 2015).

2.3.4 Verbe-Interessenstest

Im Kontext der Berufs- und Studienwahl ist das Erfassen von beruflichen Interessen ein zentrales Element. Dabei greifen die herkömmlichen Verfahren oftmals auf deskriptive Items zurück, welche berufstypische Tätigkeiten detailliert beschreiben um dann von Testpersonen bewertet zu werden (vgl. AIST-R; Bergmann & Eder, 2004; Explorix; Jörin et al., 2003). Diese Art von Items bringen aber einige Probleme mit sich. Sie sind oftmals kontextgebunden, kognitiv aufwändig und zeitintensiv. Um diesen Herausforderungen zu begegnen, entwickelten Wetzel et al. (2012) den Verb-Interessenstest (VIT), welcher zur Erfassung beruflicher Interessen kontextlose Tätigkeitsverben verwendet. Das Ziel war es einen Interessenstest zu schaffen, der zeitsparend, kontextunabhängig und über alle Bildungsniveaus einsetzbar ist. Die theoretische Grundlage des VITs bildet das RIASEC-Modell von Holland (1997). Die Verben basieren dabei auf den Dimensionen aus dem Modell. So wird beispielsweise das Verb *hämmern* der Dimension Realistic und das Verb *singen* der Dimension Artistic zugeordnet (Wetzel et al., 2012). Ein weiterer Nachteil vieler

klassischer Verfahren ist, dass diese oft geschlechterspezifische Unterschiede reproduzieren. So präferieren Frauen die Arbeit mit Menschen und besitzen eine höhere Ausprägung in den Interessensdimensionen Social, Artistic und Conventional, während Männer die Arbeit mit Gegenständen bevorzugen und eine stärkere Ausprägung in den Interessensdimensionen Realistic und Investigative aufweisen (Su, Rounds & Armstrong, 2009). Es existieren generell zwei verschiedene Ansichten, ob geschlechterspezifische Unterschiede in Interessenstests explizit behandelt und reduziert werden sollen. Einerseits wird davon ausgegangen, dass Testverfahren nur wirkliche Unterschiede in zugrundeliegenden Konstrukten aufweisen und die aktive Reduktion dieser Unterschiede die Validität der Tests verringern könnte (Gottfredson & Holland, 1978, zitiert nach Wetzel et al., 2012, S. 88). Andererseits wird argumentiert, dass das Ziel von Interessenstests die Exploration möglichst vieler unterschiedlicher Karrierepfade sein sollte, ohne dabei durch geschlechterspezifische Stereotype eingeschränkt zu werden (Prediger & Cole, 1975). Da die Items des VITs Verben sind, welchen kein Kontext gegeben wurde, sind diese weniger anfällig für geschlechterspezifische Stereotype. Beispielsweise ist das Verb *pflanzen* ein geschlechterneutrales Verb. Wird diesem aber nun ein Kontext gegeben, wie beispielsweise einen Blumengarten pflanzen oder Weizen auf dem Acker pflanzen, könnte dies Geschlechterstereotypen hervorrufen.

Zur Konstruktion des Itempools des VITs wurden gebräuchliche aktivitätsbezogene Verben mithilfe eines lexikalischen Ansatzes in einem deutschen Wörterbuch gesucht. Daraus resultierte eine Liste mit insgesamt 2'227 Verben, welchen von zwei Fachpersonen, im Bereich der Interessensforschung, jeweils eine der sechs RIASEC-Dimensionen zugeordnet wurden. Neben der Zuordnung von RIASEC-Dimensionen wurden mehrdeutige Verben identifiziert und aus der Liste entfernt. Aus der ursprünglichen Liste von Verben konnten nur 30% eindeutig einer Dimension zugeordnet werden. Insgesamt stimmte die Klassifizierung durch die Fachpersonen bei 222 Verben komplett überein. Diese Verben waren nicht

gleichmässig über die sechs Dimensionen verteilt und die Fachpersonen sollten daraus für jede der Dimensionen ungefähr 30 Verben identifizieren, welche diese am besten repräsentieren. Daraus resultierte eine Liste mit 143 Verben, welche für eine Pilotstudie ausgewählt wurden. In der Pilotstudie wurden Studierende und Mitarbeitende einer Universität ($N = 21$) gebeten, die Tätigkeitsverben auf einer 5-stufigen Likert Skala zu bewerten, anhand daran, wie gerne sie diese Tätigkeiten ausüben würden. Zusätzlich wurden die Teilnehmenden befragt, ob die Items eindeutig verständlich wären. Insgesamt wurden 31 Verben als mehrdeutig eingeschätzt. Diese Verben wurden gemeinsam mit Items, welche entweder einen Boden- oder Deckeneffekt im Antwortverhalten aufwiesen, aus dem Itempool ausgeschlossen. Von den 143 Verben blieben schlussendlich noch 113 übrig (Wetzel et al., 2012).

Die 113 Verben aus der Pilotstudie wurden im von Hell, Pässler & Schuler (2009) entwickelten Test *was-studiere.ich.de* (WSI-DE) als Zusatzmodul integriert. Beim WSI-DE handelt es sich um einen online frei zugänglichen Selbsttest, zur systematischen Erfassung von studien- und berufsrelevanter Interessen. Auch dieser Test basiert auf der Interessentheorie von Holland dem RIASEC-Modell und nach Absolvieren erhalten Teilnehmende ein detailliertes Fähigkeiten- und Interessensprofil (Hell et al., 2009). Zur Überprüfung der konvergenten Validität wurde zusätzlich der AIST-R (Bergmann & Eder, 2005) als Vergleichsinstrument miteingefügt. Wie in Kapitel 2.5.2 handelt es sich beim AIST-R um einen Berufsinteressenstest, basierend auf der Theorie von Holland. Schlussendlich wurden Teilnehmende nach ihrer Zufriedenheit mit ihrem gewählten Studienfach befragt und um eine Einschätzung gebeten, wie gut die Passung zwischen ihrem Studienfach und ihren Interessen ist. Aus einem Sample von 1'186 Testpersonen, welche den VIT ausfüllten, füllten 158 den VIT, den AIST-R und die Befragung zu ihrem Studienfach aus. Die Daten wurden mittels Hauptkomponentenanalyse ausgewertet. Dabei wurden Items, welche ihre höchste Faktorladung in einer anderen RIASEC-Dimension als der von den Fachpersonen

zugeschriebenen aufwiesen, entfernt. Items mit einer Faktorladung unter .40 wurden ebenfalls entfernt und der Itempool schrumpfte weiter auf 101 Verben. Auf diesen wurde eine weitere Hauptkomponentenanalyse angewendet und Items mit einer Querladung über .30 wurden entfernt. Nicht ausgeschlossen wurden diese Items, wenn die Querladung in einer benachbarten Dimension auftrat, da benachbarte Dimensionen ähnliche Interessen repräsentieren. Letztlich wurden weitere Items entfernt, da diese inhaltlich zu wenig trennscharf waren. Der finale Itempool umfasste noch 84 Verben, 13 aus der Dimension Realistic, 14 aus der Dimension Investigative, 9 aus der Dimension Artistic, 18 aus der Dimension Social, 18 aus der Dimension Enterprising und 12 aus der Dimension Conventional (Wetzel et al., 2012).

Um den VIT zu erstellen wurden drei verschiedene Strategien genutzt und kontrolliert, inwiefern sich die Versionen unterscheiden. Es wurde ein Test basierend auf Classical Test Theory (CTT) erstellt (VIT-Classic). Eine weitere Version wurde ebenfalls anhand der CTT erstellt, jedoch wurde versucht, die Geschlechterunterschiede in den Mittelwerten zu verringern (VIT-Balanced). Die dritte Version wurde auf Basis der Item Response Theory (IRT) erstellt und zusätzlich wurde beachtet, ob die Items für Männer und Frauen mit vergleichbarer Interessensausprägung unterschiedlich funktionieren (VIT-IRT) (Wetzel et al., 2012). Bei der Herangehensweise der CTT wird davon ausgegangen, dass die Testergebnisse einer Person sich aus dem echten Wert und einem Messfehler zusammensetzen. Die Auswahl der Items orientiert sich dabei primär an der Trennschärfe eines Items und zielt auf eine Maximierung der Reliabilität ab. Diese gibt an wie stark ein beobachteter Wert mit dem echten Wert zusammenhängt und liefert somit Hinweise darauf, wie gut ein latentes Merkmal erfasst wird (Reeve & Mâsse, 2004). Im Gegensatz dazu wird bei der IRT das Antwortverhalten auf der Itemebene analysiert (Embretson & Reise, 2000). Dadurch lässt sich erkennen, inwieweit die Antwortkategorien eines Items zwischen Testpersonen mit unterschiedlichen Ausprägungen eines latenten Merkmals differenzieren und wie gut die

Vorhersagekraft des Modells in Bezug auf die beobachteten Werte ist – also mit welcher Wahrscheinlichkeit eine Testperson mit einer spezifischen latenten Merkmalsausprägung eine bestimmte Antwortkategorie auswählt (Schmidt-Atzert, Krumm & Amelang, 2021). Es gibt dabei verschiedene Modelle um Charakteristiken von Items zu latenten Merkmalen zuzuordnen. Zu den Bekanntesten gehören das ein-, zwei- und drei-parametrische logistische Modell für dichotome Daten (Samejima, 1969, 1996, zitiert nach Wetzel et al., 2012, S. 89) sowie das Partial Credit Model (PCM; Masters, 1982, zitiert nach Wetzel et al., 2012, S. 89) für kategoriale Daten mit mehr als zwei Ausprägungen. Das Ziel bei jeder dieser Konstruktionsstrategien war, dass schlussendlich für jede RIASEC-Dimension acht Items übrig bleiben, ausser für die Dimension Artistic, für die noch fünf Items übrig bleiben sollten (Wetzel et al., 2012).

Zur Konstruktion des VIT-Classic wurden zuerst die Itemtrennschärfen und -schwierigkeiten berücksichtigt. Für den Fall, dass danach mehr als acht Items je Dimension verblieben, wurde zusätzlich die Inter-Item-Korrelation miteinbezogen. Falls nach diesem Schritt noch immer zu viele Items für die Dimensionen vorhanden waren, wurde darüber hinaus noch die Item-zu-Skala-Korrelation herangezogen (Wetzel et al., 2012).

Der VIT-Balanced folgte einer ähnlichen Konstruktionsstrategie wie der VIT-Classic. Der zentrale Unterschied war jedoch, dass die Items anhand der itembezogenen Mittelwertsunterschiede zwischen Frauen und Männern und nicht anhand der Trennschärfe aussortiert wurden. Dabei wurden jeweils die Items mit den geringsten Mittelwertsunterschieden behalten (Wetzel et al., 2012).

Um den VIT-IRT zu erarbeiten, wurde das PCM verwendet. Ausserdem wurde eine Differential Item Functioning Analyse (DIF) durchgeführt. Diese schätzt die Ausprägung latenter Merkmale in Stichprobengruppen, wie beispielsweise Frauen und Männer. Im Falle des VIT-IRT wurde geprüft, wie wahrscheinlich eine bestimmte Antwortkategorie von Frauen und Männern mit ähnlichen latenten Merkmalsausprägungen ausgewählt wird (Osterlind &

Everson, 2009). Wenn der DIF-Kontrast dabei grösser als .43 Logits war, wurden die Items ausgeschlossen. Für den Fall, dass dies nicht ausreichte um die Items entsprechend zu reduzieren, wurden die Schwellen weiter angepasst (Wetzel et al., 2012).

Aus den 84 Verben des Itempools blieben noch 72 Verben übrig die mindestens in einem der drei Testversionen Anwendung fanden (siehe Tabelle 1).

Tabelle 1

Verben aller VIT-Versionen mit zugehörigen RIASEC-Dimensionen (Wetzel et al., 2012).

| RIASEC-Dimension | Verben | |
|------------------|---------------|-----------------------|
| Realistic | aufbauen | hämmern |
| | schnitzen | pflanzen |
| | zerlegen | reparieren |
| | umgraben | schrauben |
| | bohren | tapezieren |
| | kleben | |
| Investigative | vermuten | rätseln |
| | folgern | erkennen |
| | entdecken | auskundschaften |
| | finden | testen |
| | erfinden | ausdenken |
| | nachforschen | probieren |
| | beobachten | |
| Artistic | gestalten | reimen |
| | malen | singen |
| | ausstellen | übersetzen |
| | verfilmen | dichten |
| Social | begleiten | helfen |
| | betreuen | belehren |
| | beistehen | lehren |
| | trösten | zuhören |
| | beraten | behüten |
| | ausbilden | erziehen |
| | einfühlen | jdm. etwas beibringen |
| beschützen | | |
| Enterprising | erkämpfen | investieren |
| | handeln | leiten |
| | durchgreifen | verhandeln |
| | überzeugen | überreden |
| | führen | anwerben |
| | durchsetzen | auffordern |
| | beeinflussen | fordern |
| Conventional | verwalten | nummerieren |
| | beantragen | aufzählen |
| | kalkulieren | ordnen |
| | kontrollieren | aussortieren |
| | zählen | überweisen |
| | einordnen | |

Die drei Versionen des VITs überlappen sich insbesondere in den Dimensionen Realistic und Conventional¹. Um die Kriteriumsvalidität zu untersuchen wurde die wahrgenommene Passung zwischen dem Studienfach und den eigenen Interessen und die eigene Zufriedenheit mit dem Studienfach als Kriterium definiert. Es konnten keine systematischen Unterschiede festgestellt werden und die kriterienbezogene Validität konnte nur in bestimmten Teilstichproben bestätigt werden. Trotz Unterschiede innerhalb der Teilstichproben von Frauen und Männern, konnte für alle drei Versionen des VITs die Konstruktvalidität bestätigt werden. Bei Männern wurden höhere Korrelation für die Dimensionen Realistic und Enterprising und bei Frauen für die Dimensionen Investigative und Social festgestellt. Die Limitationen der Studie bestanden einerseits in der kleinen Stichprobe ($N = 158$) zur Überprüfung der Konstruktvalidität. Andererseits wurde kritisiert, dass zur Erfassung von Berufsinteressen kontextlose Verben teilweise zu unspezifisch sind (Wetzel et al., 2012).

3 Methodisches Vorgehen

In diesem Kapitel wird das methodische Vorgehen zur Datenaufbereitung und -auswertung näher erläutert. Sämtliche Datenverarbeitung wurde mittels Python in der Codierumgebung *Jupyter Notebook* ausgeführt. Das Kapitel ist aufgeteilt in die Auswahl eines Word-Embeddings, die Aufbereitung der relevanten Daten und abschliessend die Hauptkomponentenanalyse.

3.1 Auswahl eines Word-Embeddings

Um die Verben des VITs im Rahmen eines Word-Embeddings verarbeiten zu können, wurde ein deutsches Modell benötigt. Da das Trainieren eines eigenen Modells aber den

¹ Die Zusammensetzung der drei Versionen des VITs ist in Anhang A ersichtlich.

Rahmen dieser Arbeit überschritten hätte, wurde nach vortrainierte Word-Embeddings recherchiert. Es konnten insgesamt vier, für das Vorhaben potenziell geeignete, Modelle gefunden werden. Darunter basierten zwei Word-Embeddings auf dem Word2Vec-Algorithmus (Mikolov, Sutskever, et al., 2013), eines auf dem GloVe-Algorithmus (Pennington et al., 2014) und das letzte auf dem fastText-Algorithmus (Joulin, et al., 2016). In dieser Arbeit wurden den unterschiedlichen Modellen eigene Namen zugewiesen. Eines der Word2Vec-Modelle wurde im Rahmen einer Bachelorarbeit trainiert (Müller, 2015) und für diese Arbeit „BA-Word2Vec“ genannt. Das zweite Word2Vec-Modell stammte vom deutschen Unternehmen deepset (Deepset, 2018) und trug in dieser Arbeit den Namen „Deepset-Word2Vec“. Auch das GloVe-Modell stammte von deepset (Deepset, 2023) und wurde in dieser Arbeit als „Deepset-GloVe“ bezeichnet. Das fastText-Modell wurde im Rahmen einer Studie zum Training von Word-Embeddings in 157 verschiedenen Sprachen trainiert (Grave, Bojanowski, Gupta, Joulin & Mikolov, 2018) und wurde für diese Arbeit „WE-fastText“ genannt. Die Word-Embeddings wurden als Parquet-Dateien abgespeichert, um diese in *Jupyter Notebook* weiterverarbeiten zu können.

Um ein für das Vorhaben geeignetes Word-Embedding auszuwählen, mussten alle Modelle evaluiert werden. Zur Evaluation wurden die Word-Embeddings mittels Pandas als Dataframes in *Jupyter Notebook* geladen. Das Vorgehen der Evaluation erfolgte nach einem Ausschlussverfahren. Wenn in einem der Evaluationsschritte ein Word-Embedding im Vergleich zu den restlichen auffällig schlechter abschnitt, wurde dieses ausgeschlossen und die weitere Evaluationsschritte erfolgten anschliessend mit den verbliebenen Modellen. Wie sich das Ausschliessen eines Modelles rechtfertigte, wird im Kapitel 4.1, den Ergebnissen der Evaluation, beschrieben.

3.1.1 *Abgleich von Word-Embedding und VIT*

Neben den Evaluationsmethoden zur Leistung von Word-Embeddings, musste zusätzlich eine für das Vorhaben essenzielle Voraussetzung überprüft werden. Damit ein Modell für die Arbeit in Betracht gezogen werden konnte, mussten die 72 Verben des VITs darin enthalten sein. Da es sich dabei um eine zentrale Voraussetzung handelte, wurde dies vor der eigentlichen Evaluation der Modelle geprüft. Für diese Überprüfung wurden die Verben des VITs zur Verarbeitung als Textdatei abgespeichert und in *Jupyter Notebook* geladen. Danach erfolgte ein Abgleich der Verben aus dem VIT mit den in den Word-Embeddings enthaltenen Wörtern. Das Verb *jdm. etwas beibringen* wurde zusätzlich noch in der ausgeschriebenen Form *jemandem etwas beibringen* geprüft.

3.1.2 *Intrinsische Evaluation*

Die Evaluation von Word-Embeddings teilt sich klassischerweise in die intrinsische und extrinsische Evaluation auf (Ascari, Giabelli, Malandri, Mercurio & Mezzanzanica, 2024). Die intrinsische Evaluation konzentriert sich auf die Qualität des Modells im Hinblick auf syntaktische oder semantische Beziehungen der Worte (Wang, Wang, Chen, Wang & Kuo, 2019). Bei der extrinsischen Evaluation wird geprüft, wie gut ein Modell im Kontext von konkreten Aufgaben, wie beispielsweise einer Textklassifikation, funktioniert (Camacho-Collados & Pilehvar, 2018). Für die vorliegende Arbeit war insbesondere die intrinsische Evaluation von Interesse, da die semantischen und syntaktischen Beziehungen der Verben aus dem VIT untersucht wurden. Aus diesem Grund wurde auf eine extrinsische Evaluation der Word-Embeddings verzichtet. Zur intrinsischen Evaluation empfehlen Ascari et al. (2024) die Vorgehensweise VEC2BEST, welche eine intrinsische Gesamtevaluation darstellt. VEC2BEST setzt sich aus den vier intrinsischen Evaluationen *Word similarity*, *Word analogy*, *Concept Categorization* und *Outlier Detection* zusammen, welche nach Wang et al. (2019) auch zu den effektivsten gehören. Aufgrund der Evaluation nach dem

Ausschlussverfahren wurde VEC2BEST nicht als ganzheitliche Methode angewandt, sondern beschränkte sich auf die beiden intrinsischen Evaluationen *Word similarity* und *Word analogy*.

3.1.2.1 Word Similarity

Word similarity (Wortähnlichkeitsbewertung) ist einer der verbreitetsten intrinsischen Evaluationen. Das Ziel dieser Evaluation ist es zu prüfen, wie gut ein Word-Embedding Wortähnlichkeiten aus menschlicher Sicht abbilden kann. Das Word-Embedding berechnet die Kosinuns-Ähnlichkeit für ein Wortpaar. Dieser vom Modell berechnete Wert wird dann mit den von Menschen vergebenen Werten mittels Spearman-Korrelation verglichen (Faruqui, Tsvetkov, Rastogi & Dyer, 2016). Je höher der Korrelationskoeffizient, desto näher sind die beiden Ähnlichkeitswerte aneinander und desto besser schneidet ein Modell bei dieser Evaluation ab. (Ascari et al., 2024).

Da deutsche Word-Embeddings evaluiert wurden, wurde die deutsche Version des WordSim-353 benutzt. Dieser enthält 353 deutsche Wortpaare und Ähnlichkeitsbewertungen (Leviant & Reichart, 2015). Zuerst wurde die Kosinus-Ähnlichkeit für die Wortvektoren dieser Wortpaare aus den jeweiligen Word-Embeddings berechnet (Wang et al., 2019). In einem zweiten Schritt wurde dann eine Spearman-Korrelation zwischen den Ähnlichkeitswerten des Modells und der von Menschen vergebenen Ähnlichkeitswerten aus dem WordSim-353 berechnet.

3.1.2.2 Word analogy

Im Falle der Evaluation *Word analogy*, soll ein Word-Embedding Wortbeziehungen aufdecken (Wang et al., 2019). Es werden zwei Wortpaare gewählt, welche eine Art von Beziehung zueinander haben. Ein Beispiel dafür wären die Wortpaare „Frau; Mann“ und „Königin; König“, wobei eines der Worte nicht vorgegeben wird. Es könnte beispielsweise

das Wort „König“ nicht vorgegeben werden und die Aufgabe des Modelles wäre dann, die Frage zu beantworten „Frau ist zu Mann wie Königin zu ...?“ (Levy & Goldberg, 2014). Die Berechnung des passendsten Wortes erfolgt durch Vektorarithmetik. Im genannten Beispiel würde folgende Berechnung durchgeführt werden: $Vector(Königin) - Vector(Frau) + Vector(Mann) = Vector(?)$ (Mikolov, Yih & Zweig, 2013). Um diese Evaluation durchzuführen wurde ein Datensatz mit deutschen Wortanalogien genutzt. In diesem Datensatz waren insgesamt 18‘552 deutsche Wortanalogien enthalten. Es handelte sich dabei um die deutsche Übersetzung des Analogiedatensatzes von Google (Köper, Scheible & Schulte im Walde, 2015). Um zu überprüfen wie die Leistung des Modelles im Lösen von Wortanalogien war, wurde die Accuracy über alle Vorhersagen berechnet.

Dies beendete die Evaluation der Word-Embeddings und die nachfolgenden Schritte wurden jeweils mit dem Word-Embedding „WE-fastText“ durchgeführt. Die Ergebnisse der Evaluationen und des Auswahlverfahrens können aus dem Kapitel 4.1 entnommen werden.

3.2 Vorbereitung der Daten für die Hauptkomponentenanalyse

Um anhand der Verben des VITs eine Hauptkomponentenanalyse durchführen zu können, mussten einige Vorbereitungen vorgenommen werden. Eine Hauptkomponentenanalyse mit einem Datensatz der weniger Datenpunkte als Dimensionen ($n < p$) enthält, wird als problematisch angesehen und könnte die Analyse einschränken (Jolliffe & Cadima, 2016). Da der VIT 72 Verben enthält, aber das Word-Embedding insgesamt 300 Dimensionen umfasst, musste zuerst eine Dimensionsreduktion auf mindestens 72 Dimensionen durchgeführt werden. Diese Dimensionsreduktion wurde nicht mit den Vektoren aller im Word-Embedding repräsentierten Wörter durchgeführt, sondern mit einer möglichst umfangreichen Anzahl an Verben. Der Grund dafür war, dass sich die Dimensionen des reduzierten Vektorraums möglichst nur auf Verben beziehen sollten. In einem Online-Lexikon (Göbel, 2018) konnte eine umfangreiche Liste an Verben gefunden werden. Diese

umfasste insgesamt 16'625 Verben, welche in Form einer Textdatei für die weitere Verarbeitung abgespeichert wurde. Für die Verben des VITs lag bereits eine Textdatei aus dem vorangegangenen Schritt zur Auswahl eines Word-Embeddings vor. Die beiden Textdateien wurden gemeinsam mit dem Word-Embedding in ein Jupyter Notebook geladen.

Das Word-Embedding wurde bereinigt, indem Duplikate, also komplett identische Zeilen im Word-Embedding, entfernt wurden. Anschliessend wurden die beiden Verbenlisten bereinigt, indem überflüssige Leerzeichen entfernt, Umlaute definiert und die Verben in Kleinbuchstaben umgewandelt wurden. Um zu garantieren, dass die Verben des VITs in der umfangreichen Verbenliste enthalten sind, wurden diese ebenfalls miteinander abgeglichen.

Dann konnte die Dimensionsreduktion durchgeführt werden. Eine weitverbreitete Methode zur Reduktion von Dimensionen in Word-Embeddings ist die Hauptkomponentenanalyse (Treisman, Mughaz, Stulman & Dvir, 2022; Xue, Wang, Wei & Kuo, 2024). Zur Dimensionsreduktion wurden die Vektoren der Verben der umfangreichen Verbenliste aus dem Word-Embedding extrahiert. Anhand dieser Verben wurden die 300 Dimensionen des Word-Embeddings mittels Hauptkomponentenanalyse auf die maximal mögliche Anzahl an Dimensionen reduziert. Diese Anzahl an Dimensionen wurde gewählt, um eine möglichst hohe erklärte Gesamtvarianz zu erreichen.

Anschliessend wurden die Vektoren der Verben des VITs aus dem reduzierten Dimensionsraum extrahiert und mittels Pandas in ein Dataframe abgespeichert. Dies beendete die Aufbereitung der Daten und die Grundlage für die weitere Hauptkomponentenanalyse war gegeben.

3.3 Hauptkomponentenanalyse

Um zu überprüfen, ob sich die Struktur des RIASEC-Modells anhand der Repräsentationen der Verben im Word-Embedding widerspiegeln, wurde eine Hauptkomponentenanalyse durchgeführt. Die Hauptkomponentenanalyse ist nicht nur eine

gängige Methode zur Reduktion von Dimensionen in Word-Embeddings (Treisman et al., 2022; Xue et al., 2024), sondern eignet sich auch dazu latente Strukturen in diesen zu entdecken (vgl. Bolukbasi, Chang, Zou, Saligrama & Kalai, 2016; Vargas & Cotterell, 2020; Zhao et al., 2019). Um eine Hauptkomponentenanalyse durchführen zu können, gibt es einige allgemeine Voraussetzungen, welche erfüllt sein müssen. Zunächst sollte es einen linearen Zusammenhang zwischen den Variablen geben (Bakker, 2024). Um dies zu prüfen, wurde für die Dimensionen des reduzierten Word-Embeddings eine Korrelationsmatrix erstellt und zusätzlich die durchschnittliche absolute Korrelation berechnet. Die Korrelationsmatrix² sowie die durchschnittliche absolute Korrelation ($r = .12$) wiesen auf einen schwachen, aber vorhandenen linearen Zusammenhang hin. Eine weitere Voraussetzung ist, dass die Variablen kontinuierlich sind (Bakker, 2024). Da es sich bei Word-Embeddings um Modelle kontinuierlicher Wortrepräsentationen handelt, war diese Voraussetzung gegeben (Naseem, Razzak, Khan & Prasad, 2021). Weil die Hauptkomponentenanalyse empfindlich gegenüber Ausreißern ist (Neumayer, Nimmer, Setzer & Steidl, 2020), wurden die Daten vor der Extraktion der Hauptkomponenten mittels z-Wert-Verfahren überprüft. Zur Identifikation potenzieller Ausreisser wurde das gängige Kriterium verwendet, nach dem z-Werte, die im Absolutwert grösser als 3 sind ($|z| > 3$), als Ausreisser gelten (Iglewicz & Hoaglin, 1993). Dies identifizierte insgesamt 17 Verben als potenzielle Ausreisser (siehe Tabelle 2).

² Die Korrelationsmatrix des reduzierten Word-Embeddings ist als Heatmap in Anhang B ersichtlich.

Tabelle 2

Als Ausreisser identifizierte Verben mit zugeordneter RIASEC-Dimension.

| Ausreisser-Verb | RIASEC-Dimension |
|-----------------|------------------|
| bohren | Realistic |
| hämmern | |
| kleben | |
| tapezieren | |
| umgraben | |
| folgern | Investigative |
| rätseln | |
| vermuten | |
| dichten | Artistic |
| malen | |
| reimen | |
| singen | |
| behüten | Social |
| einfühlen | |
| trösten | |
| zuhören | |
| anwerben | Enterprising |

Die Ausreisser-Verben wurden nicht entfernt, da es sich dabei um eine grosse Anzahl handelte. Durch das Entfernen würden ansonsten in einzelnen RIASEC-Dimensionen nur noch eine geringe Anzahl an Verben verbleiben. Ein weiterer Grund diese Verben nicht zu entfernen war, dass es sich dabei auch um semantisch besonders markante Verben handeln könnte.

Bei einer Hauptkomponentenanalyse ist zudem eine Mittelwertzentrierung nötig, sodass die erste Hauptkomponente die Richtung der grössten Varianz beschreibt. Wird diese Zentrierung nicht gemacht, kann diese stattdessen in die Richtung des Mittelwerts zeigen, wodurch die Analyse verfälscht wird (Roch, 2025). Zur Durchführung der Hauptkomponentenanalyse wurde die PCA-Funktion von *sklearn* in *Jupyter Notebook* verwendet. Bei dieser Funktion wird die Mittelwertzentrierung automatisch vorgenommen

(scikit-learn, n.d.). Zur Sicherheit wurde dies geprüft, indem nochmals manuell eine Mittelwertzentrierung durchgeführt wurde. Die Ergebnisse der darauffolgenden Hauptkomponentenanalyse waren dabei identisch mit den Ergebnissen ohne manuelle Mittelwertzentrierung. Dies bestätigte die Annahme zur PCA-Funktion von *sklearn*.

Eine Standardisierung der Daten ist bei einer Hauptkomponentenanalyse besonders dann empfohlen, wenn diese unterschiedliche Einheiten oder Skalen aufweisen (Jolliffe & Cadima, 2016). Dies trifft nicht auf die Vektoren eines Word-Embeddings zu, jedoch können die Vektoren von unterschiedlicher Länge sein. Eine Standardisierung der Vektoren wird jedoch nicht empfohlen, da auch die Länge von Vektoren wichtige semantische Informationen enthalten kann, welche bei einer Standardisierung verloren gehen würden (Schakel & Wilson, 2015). Aus diesem Grund wurde auf eine Standardisierung der Daten verzichtet.

Da das RIASEC-Modell aus sechs Dimensionen besteht und untersucht wurde, ob sich diese Dimensionen anhand der Verben des VITs im Word-Embedding widerspiegeln, wurden die ersten sechs Faktoren extrahiert. Zusätzlich zu dieser theoretischen Annahme zur Extraktion wurde ein Screeplot (siehe Abbildung 5) erstellt. Auch anhand des Screeplots schien die Extraktion von sechs Faktoren als angemessen.

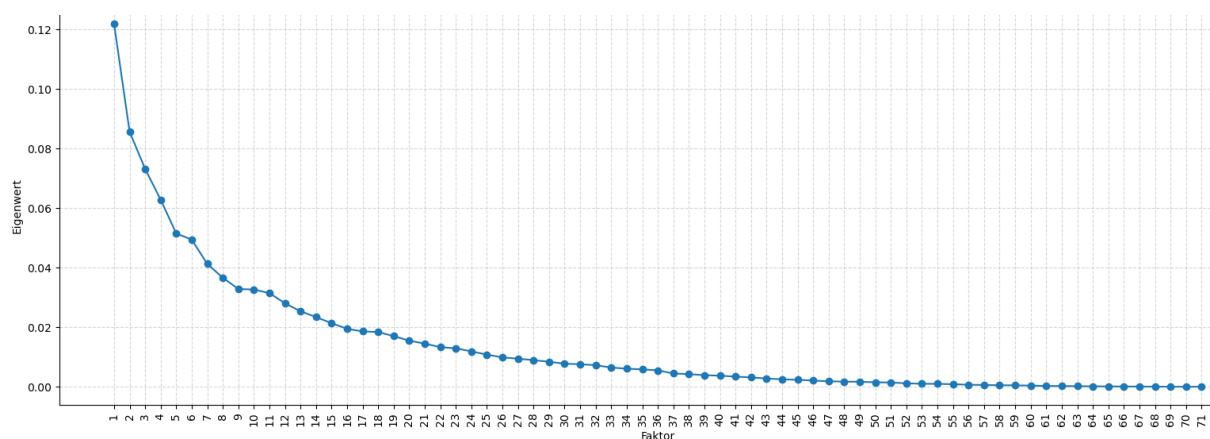


Abbildung 5. Screeplot zur Bestätigung der Faktorenanzahl bei der Hauptkomponentenanalyse

Um das Ergebnis der Hauptkomponentenanalyse einfacher interpretieren zu können, wurde die extrahierte 6-Faktoren-Lösung einer Oblimin-Rotation unterzogen. Da analog zum RIASEC-Modell von einem Zusammenhang zwischen den Faktoren ausgegangen wurde, wurde eine oblique Rotation gewählt (Field, 2018). Um die Ergebnisse der Hauptkomponentenanalyse entlang des RIASEC-Modells inhaltlich besser interpretieren zu können, wurde zusätzlich für jeden Faktor der Mittelwert der Faktorladungen je RIASEC-Dimension berechnet.

4 Ergebnisse

Im nachfolgenden Kapitel werden die Ergebnisse der zuvor beschriebenen Analysen dargestellt. Nach der Schilderung der Resultate aus der Auswahl eines Word-Embeddings und der Vorbereitung der Daten, wird auf die Ergebnisse der Hauptkomponentenanalyse eingegangen.

4.1 Ergebnisse der Auswahl eines Word-Embeddings

In diesem Kapitel werden die Ergebnisse der Auswahl eines Word-Embeddings beschrieben. Die Auswahl erfolgte nach dem Ausschlussverfahren. In den folgenden Unterkapiteln wird beschrieben wie die einzelnen Word-Embeddings im jeweiligen Schritt abgeschnitten und wann ein Word-Embedding aus welchem Grund ausgeschlossen wurde.

4.1.1 Ergebnisse des Abgleichs von VIT und Word-Embeddings

Die erste zu prüfende Voraussetzung war, ob die Verben des VITs in den vorliegenden Word-Embeddings enthalten waren. Die Ergebnisse aus diesem Abgleich sind in Tabelle 3 ersichtlich.

Tabelle 3

Anzahl enthaltener Verben aus dem VIT in den Word-Embeddings.

| BA-Word2Vec | Deepset-Word2Vec | Deepset-GloVe | WE-fastText |
|--------------------|-------------------------|----------------------|--------------------|
| 56 von 72 Verben | 56 von 72 Verben | 71 von 72 Verben | 71 von 72 Verben |

In den Word-Embeddings BA-Word2Vec und Deepset-Word2Vec fehlten insgesamt 16 Verben, es handelte sich dabei in beiden Word-Embeddings um dieselben Verben. In den Word-Embeddings Deepset-GloVe und WE-fastText fehlte jeweils nur ein Verb³. In allen Word-Embeddings fehlte das Verb *jdm. etwas beibringen*, weshalb dieses Verb zusätzlich in ausgeschriebener Form geprüft wurde. Auch in ausgeschriebener Form war dieses Verb in keinem der Word-Embeddings vorhanden. Aus der Liste der Verben des VITs, war dies das einzige Verb, welches von Pronomen begleitet wird. Dies könnte erklären weshalb das Verb in allen Modellen fehlte, da wie in Kapitel 2.2.1 erwähnt Word-Embeddings nur einzelne Wörter repräsentieren. Aufgrund des hohen Anteils an fehlenden Verben in den Modellen BA-Word2Vec und Deepset-Word2Vec wurden diese von der weiteren Evaluation ausgeschlossen.

4.1.2 Ergebnisse der intrinsischen Evaluation der Word-Embeddings

Die intrinsischen Evaluationen *Word similarity* und *Word analogy* erfolgten mit den beiden verbliebenen Word-Embeddings: Deepset-GloVe und WE-fastText. Die Ergebnisse dieser Evaluationen sind in Tabelle 4 ersichtlich.

Tabelle 4

Ergebnisse der intrinsischen Evaluation.

| | Word similarity (r) | Word analogy (Accuracy) |
|----------------------|---|--------------------------------|
| Deepset-GloVe | .450 | 0.56% |
| WE-fastText | .498 | 61% |

³ Die Liste der in den Word-Embeddings fehlenden Verben ist in Anhang C ersichtlich.

Die Ergebnisse der Evaluation *Word similarity* ergab für beide Word-Embeddings eine mittlere Korrelation, die knapp nicht einer hohen Korrelation entspricht (Kuckartz, Rädiker, Ebert & Schehl, 2013). Das Modell WE-fastText ($r = .498$) schnitt dabei etwas besser ab als das Modell Deepset-GloVe ($r = .450$). Da der Unterschied nur sehr gering war, wurde nach dieser Evaluation keines der Word-Embeddings ausgeschlossen. Ein weitaus deutlicher Unterschied zeigte sich bei der Evaluation *Word analogy*. Während WE-fastText eine Accuracy von 61% erreichen konnte, erreichte Deepset-GloVe nur 0.56%. Dieses Ergebnis wurde als auffällig tief wahrgenommen, was dazu führte, dass das Word-Embedding Deepset-GloVe nochmals genauer überprüft wurde. Es stellte sich heraus, dass von den 18'552 Analogien nur 5'924 getestet werden konnten. Von diesen Analogien konnte das Modell insgesamt nur 33 korrekt lösen. Diese korrekt gelösten Analogien bezogen sich alle auf syntaktische Unterschiede in den Worten wie z.B. *tanzen; tanzte* und *verkaufen; verkaufte*. Um zu kontrollieren welche Art von Worten im Deepset-GloVe repräsentiert werden, wurde eine zufällige Liste an Worten aus dem Embedding gefiltert. Dabei stellte sich heraus, dass in diesem Word-Embedding nicht nur deutsche Wörter, sondern auch Wörter in anderen Sprachen repräsentiert werden. Sogar Begriffe in kyrillischen, arabischen und chinesischen Schriftzeichen wurden gefunden. Diese Erkenntnisse waren Anlass genug, das Word-Embedding von der Auswahl auszuschliessen.

Dieser Schritt beendete die Auswahl eines Word-Embeddings und „WE-fastText“ wurde ausgewählt. Die weiteren Analyseschritte dieser Arbeit wurden mit diesem Word-Embedding durchgeführt.

4.2 Ergebnisse aus der Vorbereitung der Daten für die Hauptkomponentenanalyse

In einem ersten Schritt wurden aus dem Word-Embedding die Duplikate entfernt. Dies reduzierte die Zeilen von 4'000'000 auf 1'299'999. Der Abgleich der Verben des VITs mit

dem Inhalt der umfangreichen Verbenliste ergab, dass alle Verben, bis auf *jdm. etwas beibringen*, darin enthalten waren. Dieses Verb war auch im Word-Embedding nicht enthalten, weshalb kein Handlungsbedarf entstand. Von den 16'625 Verben der umfangreichen Verbenliste waren insgesamt 9'241 im Word-Embedding enthalten. Wie im Kapitel 4.1.1 beschrieben, konnten 71 der 72 Verben in „WE-fastText“ wiedergefunden werden. Daher wurde eine Dimensionsreduktion mittels Hauptkomponentenanalyse von 300 auf 71 Dimensionen durchgeführt. Mit dieser Anzahl an Dimensionen liess sich noch eine Gesamtvarianz von 56.73% erklären. Aus diesem reduzierten Dimensionsraum wurden die Verben des VITs und ihre Vektoren extrahiert, um damit die Hauptkomponentenanalyse durchzuführen.

4.3 Ergebnisse Hauptkomponentenanalysen

In diesem Kapitel werden die Ergebnisse der Hauptkomponentenanalyse beschrieben. Mit den sechs extrahierten Faktoren liess sich noch eine Gesamtvarianz von 44.42%⁴ erklären.

Um einen ersten Eindruck zur Übereinstimmung der Faktoren mit einzelnen RIASEC-Dimensionen zu ermöglichen, wurden die Mittelwerte der Faktorladungen pro Dimension in Tabellenform aufgeführt. Die Mittelwertstabelle bietet die Möglichkeit, einen Überblick zu erhalten, welche der Interessensdimensionen mit welchem Faktor assoziiert werden können. Im darauffolgenden Unterkapitel werden für jeden der sechs Faktoren jeweils die zehn Verben mit den stärksten positiven und negativen Faktorladungen⁵ präsentiert und beschrieben.

⁴ Die erklärte Varianz der Hauptkomponentenanalyse pro Faktor ist in Anhang D ersichtlich.

⁵ Die Vollständige Ladungsmatrix der Hauptkomponentenanalyse ist in Anhang E ersichtlich

4.3.1 Mittelwerte der Faktorladungen

Im folgenden Kapitel werden die Ergebnisse der Mittelwertberechnungen der Faktorladungen jeder RIASEC-Dimension für alle Faktoren präsentiert (siehe Tabelle 5). Die Mittelwerte der RIASEC-Dimensionen basieren jeweils auf einer leicht unterschiedlichen Anzahl an Werten ($n = 8$ bis $n = 14$). Es erfolgt zuerst eine Beschreibung des höchsten und niedrigsten Mittelwerts der einzelnen RIASEC-Dimensionen über alle Faktoren.

Anschliessend werden zusätzlich noch der höchste und niedrigste Mittelwert der RIASEC-Dimensionen innerhalb der einzelnen Faktoren beschrieben.

Tabelle 5

Mittelwerte der RIASEC-Dimensionen innerhalb der Faktoren.

| | R | I | A | S | E | C |
|-----------|---------------|---------------|---------------|---------------|---------------|---------------|
| F1 | 0.023 (0.10) | 0.049 (0.20) | -0.007 (0.07) | -0.026 (0.09) | 0.003 (0.06) | 0.031 (0.10) |
| F2 | 0.006 (0.07) | -0.009 (0.10) | 0.023 (0.08) | 0.052 (0.21) | -0.023 (0.11) | -0.014 (0.04) |
| F3 | 0.017 (0.10) | 0.003 (0.10) | 0.071 (0.15) | 0.005 (0.07) | 0.018 (0.11) | 0.037 (0.19) |
| F4 | 0.014 (0.10) | 0.030 (0.10) | -0.015 (0.11) | 0.007 (0.07) | -0.002 (0.20) | 0.001 (0.11) |
| F5 | -0.072 (0.09) | 0.041 (0.15) | -0.050 (0.15) | 0.013 (0.11) | 0.017 (0.11) | 0.018 (0.09) |
| F6 | -0.012 (0.08) | 0.048 (0.19) | 0.014 (0.11) | -0.014 (0.13) | 0.022 (0.11) | 0.006 (0.07) |

Anmerkungen. Mittelwerte (Standardabweichung in Klammern)

Die Mittelwerte der Faktorladungen zeigten, dass sich anhand dieser keine der RIASEC-Dimensionen eindeutig einem einzelnen Faktor zuweisen liess. Hinzu kam, dass in allen Fällen die Standardabweichungen die jeweiligen Mittelwerte überstiegen. Dies weist darauf hin, dass die Verben einer RIASEC-Dimension sehr unterschiedlich auf denselben Faktor luden.

Für die Dimension Realistic lag der höchste Mittelwert bei Faktor 1 ($M = 0.023$, $SD = 0.10$) und der niedrigste bei Faktor 5 ($M = -0.072$, $SD = 0.09$). Die Dimension Investigative

erreichte ebenfalls den höchsten Mittelwert bei Faktor 1 ($M = 0.049$, $SD = 0.20$) und den niedrigsten bei Faktor 2 ($M = -0.009$, $SD = 0.10$). In der Dimension Artistic zeigte sich der höchste Mittelwert bei Faktor 3 ($M = 0.071$, $SD = 0.15$) und der niedrigste bei Faktor 5 ($M = -0.050$, $SD = 0.15$). Für die Dimension Social entfiel der höchste Mittelwert auf Faktor 2 ($M = 0.052$, $SD = 0.21$) und der niedrigste auf Faktor 1 ($M = -0.026$, $SD = 0.09$). Die Dimension Enterprising erreichte den höchsten Mittelwert bei Faktor 6 ($M = 0.022$, $SD = 0.11$) und den niedrigsten bei Faktor 2 ($M = -0.023$, $SD = 0.11$). Die Dimension Conventional zeigte den höchsten Mittelwert bei Faktor 3 ($M = 0.037$, $SD = 0.19$) und den niedrigsten bei Faktor 2 ($M = -0.014$, $SD = 0.04$).

Innerhalb von Faktor 1 wies die Dimension Investigative ($M = 0.049$, $SD = 0.20$) den höchsten und die Dimension Social ($M = -0.026$, $SD = 0.09$) den niedrigsten Mittelwert auf. In Faktor 2 zeigte die Dimension Social ($M = 0.052$, $SD = 0.21$) den höchsten und die Dimension Enterprising ($M = -0.023$, $SD = 0.11$) den niedrigsten Mittelwert. Bei Faktor 3 erreichte die Dimension Artistic ($M = 0.071$, $SD = 0.15$) den höchsten und die Dimension Investigative ($M = 0.003$, $SD = 0.10$) den niedrigsten Mittelwert. Für Faktor 4 wies die Dimension Investigative ($M = 0.030$, $SD = 0.10$) den höchsten und die Dimension Artistic ($M = -0.015$, $SD = 0.11$) den niedrigsten Mittelwert auf. Innerhalb von Faktor 5 zeigte die Dimension Investigative ($M = 0.041$, $SD = 0.15$) den höchsten und die Dimension Realistic ($M = -0.072$, $SD = 0.09$) den niedrigsten Mittelwert. Für Faktor 6 erreichte die Dimension Investigative ($M = 0.048$, $SD = 0.19$) den höchsten und die Dimension Social ($M = -0.014$, $SD = 0.13$) den niedrigsten Mittelwert.

4.3.2 Beschreibung der Top-10 und Bottom-10 Faktorladungen

In diesem Kapitel erfolgen die Beschreibungen der zehn stärksten positiven und negativen Faktorladungen. Zu jedem einzelnen Faktor werden jeweils zuerst in Tabellenform die Verben, deren Faktorladungen und RIASEC-Dimensionen dargestellt und anschliessend

beschrieben. Die Tabellen wurden durch eine zusätzliche Spalte ergänzt. Diese enthält die Ausprägung auf den zugrundeliegenden Achsen des RIASEC-Modells nach Prediger (1982) und Hogan (1983, zitiert nach Sodano, 2015, S. 284) aus Kapitel 2.3.3. Den jeweiligen Ausprägungen wurden Abkürzungen zugewiesen, welche aus Tabelle 6 entnommen werden können.

Tabelle 6

Abkürzungen der Achsenausprägungen auf den zugrundeliegenden Achsen des RIASEC-Modells.

| Achse | Ausprägung | Abkürzung |
|----------------|-------------------|------------------|
| Dinge-Menschen | Dinge | DI |
| | Menschen | M |
| Daten-Ideen | Daten | DA |
| | Ideen | I |
| Conformity | Hoch | C+ |
| | Tief | C- |
| Sociability | Hoch | S+ |
| | Tief | S- |

4.3.2.1 Faktor 1

In diesem Kapitel werden die Top-10 und Bottom-10 Faktorladungen für Faktor 1 präsentiert und beschrieben. Die zehn stärksten positiven Faktorladungen sind in Tabelle 7 und die zehn stärksten negativen in Tabelle 8 dargestellt.

Tabelle 7

Top-10-Verben der 6-Faktorenlösung für Faktor 1.

| Verb | Faktorladung | RIASEC-Dimension | Achsen-Zuordnung |
|-------------|--------------|------------------|------------------|
| finden | .769 | Investigative | DI, I, S- |
| pflanzen | .245 | Realistic | DI, S- |
| beantragen | .235 | Conventional | DI, DA, C+ |
| begleiten | .178 | Social | M, S+ |
| einordnen | .168 | Conventional | DI, DA, C+ |
| schrauben | .150 | Realistic | DI, S- |
| folgern | .113 | Investigative | DI, I, S- |
| singen | .112 | Artistic | M, I, C- |
| durchsetzen | .108 | Enterprising | M, DA, S+ |
| reparieren | .086 | Realistic | DI, S- |

Unter den zehn Verben, welche am stärksten positiv auf Faktor 1 laden, liessen sich Verben aus jeder RIASEC-Dimension mindestens einmal wiederfinden. Dabei war keine der Dimensionen mit mehr als drei Verben vertreten. Damit wurde deutlich, dass Faktor 1 keine einzelne Interessensdimension des Modells eindeutig widerspiegelte.

Am häufigsten vertreten waren Verben aus der Dimension Realistic ($n = 3$), gefolgt von Verben aus den Dimensionen Investigative ($n = 2$) und Conventional ($n = 2$). Die stärkste positive Faktorladung von .769 zeigte das Verb *finden* aus der Dimension Investigative. Die weiteren positiven Faktorladungen lagen deutlich unter diesem Wert und begannen ab einer Ladung von .245.

Auf der Dinge-Menschen-Achse überwogen die Verben aus den dingeorientierten Dimensionen in ihrer Anzahl (7 zu 3) als auch hinsichtlich der stärksten positiven Faktorladung. Die Verben verteilten sich gleichmässig auf die Dimensionen der Daten-Ideen-Achse (3 zu 3), die stärkste positive Faktorladung trug dabei jedoch ein Verb aus einer ideenorientierten Dimension.

Auf der Sociability-Achse überwogen die Verben aus den Dimensionen niedriger Soziabilität sowohl in ihrer Anzahl (5 zu 2), als auch in Bezug auf die stärkste positive Faktorladung. Entlang der Conformity-Achse überwogen die Verben aus der Dimension

hoher Konformität in ihrer Anzahl (2 zu 1). Die stärkste positive Faktorladung trug ein Verb, welches aus einer Dimension stammt, die nicht auf der Conformity-Achse liegt.

Tabelle 8

Bottom-10-Verben der 6-Faktorenlösung für Faktor 1.

| Verb | Faktorladung | RIASEC-Dimension | Achsen-Zuordnung |
|-------------|---------------------|-------------------------|-------------------------|
| lehren | -.139 | Social | M, S+ |
| beraten | -.137 | Social | M, S+ |
| einfühlen | -.135 | Social | M, S+ |
| überzeugen | -.110 | Enterprising | M, DA, S+ |
| übersetzen | -.099 | Artistic | M, I, C- |
| zerlegen | -.089 | Realistic | DI, S- |
| zuhören | -.079 | Social | M, S+ |
| aufbauen | -.073 | Realistic | DI, S- |
| testen | -.066 | Investigative | DI, I, S- |
| überweisen | -.065 | Conventional | DI, DA, C+ |

Auch unter den zehn Verben, welche am stärksten negativ auf Faktor 1 luden, befand sich mindestens ein Verb aus jeder RIASEC-Dimension. Keine der Dimensionen trat dabei häufiger als viermal auf. Dies verdeutlichte, dass Faktor 1 sich auch nicht klar von einer einzelnen Interessensdimension des Modells abgrenzte.

Am häufigsten vertreten waren Verben aus der Dimension Social ($n = 4$), gefolgt von Realistic ($n = 2$). Hinsichtlich der Faktorladungen zeigte keines der Verben eine auffällig starke Negativladung im Vergleich zu den anderen Verben.

Auf der Dinge-Menschen-Achse überwogen die Verben aus den menschenorientierten Dimensionen in ihrer Anzahl (6 zu 4). Die Verteilung der Verben auf die Dimensionen der Daten-Ideen-Achse war jedoch gleichmässig (2 zu 2).

Entlang der Sociability-Achse überwogen die Verben aus den Dimensionen hoher Soziabilität in ihrer Anzahl (5 zu 3). Auf der Conformity-Achse verteilten sich die Verben gleichmässig auf die Dimensionen hoher und niedriger Konformität (1 zu 1).

4.3.2.2 Faktor 2

In diesem Kapitel werden die Top-10 und Bottom-10 Faktorladungen für Faktor 2 präsentiert und beschrieben. Die zehn stärksten positiven Faktorladungen sind in Tabelle 9 und die zehn stärksten negativen in Tabelle 10 dargestellt.

Tabelle 9

Top-10-Verben der 6-Faktorenlösung für Faktor 2.

| Verb | Faktorladung | RIASEC-Dimension | Achsen-Zuordnung |
|-------------|---------------------|-------------------------|-------------------------|
| helfen | .746 | Social | M, S+ |
| testen | .224 | Investigative | DI, I, S- |
| gestalten | .182 | Artistic | M, I, C- |
| entdecken | .136 | Investigative | DI, I, S- |
| durchsetzen | .132 | Enterprising | M, DA, S+ |
| beraten | .126 | Social | M, S+ |
| begleiten | .112 | Social | M, S+ |
| hämmern | .107 | Realistic | DI, S- |
| umgraben | .099 | Realistic | DI, S- |
| auffordern | .098 | Enterprising | M, DA, S+ |

Unter den zehn Verben, die am stärksten positiv auf den Faktor 2 luden, war mindestens ein Verb aus jeder RIASEC-Dimension, ausser aus Conventional, zu finden. Dabei war keine der Dimensionen mit mehr als drei Verben vertreten. Daran liess sich erkennen, dass Faktor 2 keine einzelne Interessensdimension des Modells eindeutig widerspiegelte.

Am häufigsten vertreten waren Verben aus der Dimension Social ($n = 3$), gefolgt von Verben aus den Dimensionen Investigative ($n = 2$), Enterprising ($n = 2$) und Realistic ($n = 2$). Die stärkste positive Faktorladung von .746 zeigte dabei das Verb *helfen* aus der Dimension Social. Die weiteren positiven Faktorladungen lagen deutlich unter diesem Wert und begannen ab einer Ladung von .224.

Die Verben aus den menschenorientierten Dimensionen der Dinge-Menschen-Achse überwogen in ihrer Anzahl (6 zu 4) und auch in Hinsicht auf die stärkste positive

Faktorladung. Auf der Daten-Ideen-Achse überwogen die Verben aus den ideenbezogenen Dimensionen in ihrer Anzahl (5 zu 3), während die stärkste positive Faktorladung aus keiner der Dimensionen auf dieser Achse stammte.

Entlang der Sociability-Achse überwogen die Verben aus den Dimensionen hoher Soziabilität in ihrer Anzahl (5 zu 4) und auch die stärkste positive Faktorladung stammte aus einer dieser Dimensionen. Nur ein Verb entfiel auf eine Dimension der Conformity-Achse. Es handelte sich dabei um ein Verb aus der Dimension niedriger Konformität, welches nicht die stärkste positive Faktorladung trug.

Tabelle 10

Bottom-10-Verben der 6-Faktorenlösung für Faktor 2.

| Verb | Faktorladung | RIASEC-Dimension | Achsen-Zuordnung |
|-------------|---------------------|-------------------------|-------------------------|
| überzeugen | -.330 | Enterprising | M, DA, S+ |
| erkennen | -.112 | Investigative | DI, I, S- |
| handeln | -.110 | Enterprising | M, DA, S+ |
| tapezieren | -.109 | Realistic | DI, S- |
| vermuten | -.105 | Investigative | DI, I, S- |
| beistehen | -.104 | Social | M, S+ |
| beobachten | -.103 | Investigative | DI, I, S- |
| aufbauen | -.096 | Realistic | DI, S- |
| einordnen | -.094 | Conventional | DI, DA, C+ |
| rätseln | -.083 | Investigative | DI, I, S- |

Unter den zehn Verben, die am stärksten negativ auf Faktor 2 luden, liessen sich Verben aus jeder RIASEC-Dimension, ausser aus Artistic, mindestens einmal wiederfinden. Darunter trat keine der Dimensionen mehr als viermal auf. Damit wurde deutlich, dass Faktor 2 sich auch nicht klar von einer einzelnen Interessensdimension des Modells abgrenzte.

Am häufigsten vertreten waren Verben aus der Dimension Investigative ($n = 4$), gefolgt von Verben aus den Dimensionen Enterprising ($n = 2$) und Realistic ($n = 2$). Die stärkste negative Ladung von $-.330$ wies dabei das Verb *überzeugen* aus der Dimension

Enterprising auf. Die weiteren negativen Ladungen lagen deutlich unter diesem Wert und begannen ab einer Ladung von -0.112 .

Auf der Dinge-Menschen-Achse überwogen die Verben aus den dingeorientierten Dimensionen klar in ihrer Anzahl (7 zu 3), jedoch nicht in Bezug auf die stärkste negative Faktorladung. Entlang der Daten-Ideen-Achse überwogen die Verben aus den ideenorientierten Dimensionen in ihrer Anzahl (4 zu 3). Die stärkste negative Ladung zeigte jedoch ein Verb aus einer datenorientierten Dimension.

Die Verben aus den Dimensionen niedriger Soziabilität, auf der Sociability-Achse, überwogen in ihrer Anzahl (6 zu 3), jedoch wies die stärkste negative Faktorladung ein Verb aus einer Dimension hoher Soziabilität auf. Auch innerhalb der negativen Faktorladungen entfiel nur ein Verb auf eine Dimension der Conformity-Achse. Es handelte sich dabei um ein Verb aus der Dimension hoher Konformität, welches nicht die stärkste negative Faktorladung trug.

4.3.2.3 Faktor 3

In diesem Kapitel werden die Top-10 und Bottom-10 Faktorladungen für Faktor 3 präsentiert und beschrieben. Die zehn stärksten positiven Faktorladungen sind in Tabelle 11 und die zehn stärksten negativen in Tabelle 12 dargestellt.

Tabelle 11

Top-10-Verben der 6-Faktorenlösung für Faktor 3.

| Verb | Faktorladung | RIASEC-Dimension | Achsen-Zuordnung |
|--------------|--------------|------------------|------------------|
| zählen | .529 | Conventional | DI, DA, C+ |
| gestalten | .359 | Artistic | M, I, C- |
| beeinflussen | .230 | Enterprising | M, DA, S+ |
| entdecken | .211 | Investigative | DI, I, S- |
| malen | .186 | Artistic | M, I, C- |
| pflanzen | .170 | Realistic | DI, S- |
| investieren | .161 | Enterprising | M, DA, S+ |
| aufbauen | .155 | Realistic | DI, S- |
| singen | .141 | Artistic | M, I, C- |
| handeln | .123 | Enterprising | M, DA, S+ |

Unter den zehn Verben, welche am stärksten positiv auf Faktor 3 luden, war mindestens ein Verb aus jeder RIASEC-Dimension, ausser aus Social, zu finden. Dabei war keine Dimension mit mehr als drei Verben vertreten. Dies verdeutlichte, dass Faktor 3 keine einzelne Interessensdimension des Modells eindeutig widerspiegelte.

Am häufigsten vertreten waren Verben aus den Dimension Artistic ($n = 3$) und Enterprising ($n = 3$), gefolgt von Verben aus der Dimension Realistic ($n = 2$). Die stärkste positive Faktorladung von .529 trug das Verb *zählen* aus der Dimension Conventional. Die weiteren positiven Faktorladungen lagen deutlich unter diesem Wert und begannen ab einer Ladung von .359.

Die Verben aus den menschenorientierten Dimensionen auf der Dinge-Menschen-Achse überwogen in ihrer Anzahl (6 zu 4), jedoch gehörte die stärkste positive Faktorladung zu einem Verb aus einer dingeorientierten Dimension. Die Verben verteilten sich gleichmässig auf die Dimensionen entlang der Daten-Ideen-Achse (4 zu 4), doch die stärkste positive Faktorladung wies ein Verb aus einer datenorientierten Dimension auf.

Auch auf der Sociability-Achse verteilten sich die Verben gleichmässig auf die Dimensionen hoher und niedriger Soziabilität (3 zu 3). Das Verb mit der stärksten positiven Faktorladung stammte aus keiner Dimension auf der Sociability-Achse. Die Verben aus der

Dimension niedriger Konformität auf der Conformity-Achse, überwogen in ihrer Anzahl (3 zu 1), jedoch trug die stärkste positive Faktorladung ein Verb aus der Dimension hoher Konformität.

Tabelle 12

Bottom-10-Verben der 6-Faktorenlösung für Faktor 3.

| Verb | Faktorladung | RIASEC-Dimension | Achsen-Zuordnung |
|--------------|---------------------|-------------------------|-------------------------|
| testen | -.253 | Investigative | DI, I, S- |
| ordnen | -.181 | Conventional | DI, DA, C+ |
| durchsetzen | -.146 | Enterprising | M, DA, S+ |
| betreuen | -.117 | Social | M, S+ |
| auffordern | -.115 | Enterprising | M, DA, S+ |
| beantragen | -.110 | Conventional | DI, DA, C+ |
| reparieren | -.106 | Realistic | DI, S- |
| führen | -.102 | Enterprising | M, DA, S+ |
| aussortieren | -.087 | Conventional | DI, DA, C+ |
| behüten | -.075 | Social | M, S+ |

Unter den zehn Verben, welche am stärksten negativ auf Faktor 3 luden, war mindestens ein Verb aus jeder RIASEC-Dimension, ausser aus Artistic, zu finden. Keine der Dimensionen trat dabei mehr als dreimal auf. Daran liess sich erkennen, dass Faktor 3 sich auch nicht klar von einer einzelnen Interessensdimension des Modells abgrenzte.

Am häufigsten vertreten waren Verben aus den Dimensionen Conventional ($n = 3$) und Enterprising ($n = 3$), gefolgt von Verben aus der Dimension Social ($n = 2$). Hinsichtlich der Faktorladungen zeigte keines der Verben eine auffällig starke Negativladung im Vergleich zu den anderen Verben.

Die Verben verteilten sich gleichmässig auf die Dimensionen der Dinge-Menschen-Achse (5 zu 5). Entlang der Daten-Ideen-Achse überwogen die Verben aus den datenorientierten Dimensionen in ihrer Anzahl (6 zu 1).

Auf der Sociability-Achse überwogen die Verben aus den Dimensionen hoher Soziabilität (5 zu 3) und auf der Conformity-Achse überwogen die Verben aus der Dimension hoher Konformität in ihrer Anzahl (3 zu 0).

4.3.2.4 Faktor 4

In diesem Kapitel werden die Top-10 und Bottom-10 Faktorladungen für Faktor 4 präsentiert und beschrieben. Die zehn stärksten positiven Faktorladungen sind in Tabelle 13 und die zehn stärksten negativen in Tabelle 14 dargestellt.

Tabelle 13

Top-10-Verben der 6-Faktorenlösung für Faktor 4.

| Verb | Faktorladung | RIASEC-Dimension | Achsen-Zuordnung |
|--------------|---------------------|-------------------------|-------------------------|
| führen | .578 | Enterprising | M, DA, S+ |
| entdecken | .291 | Investigative | DI, I, S- |
| verwalten | .180 | Conventional | DI, DA, C+ |
| kleben | .168 | Realistic | DI, S- |
| malen | .160 | Artistic | M, I, C- |
| kalkulieren | .138 | Conventional | DI, DA, C+ |
| fordern | .121 | Enterprising | M, DA, S+ |
| zuhören | .114 | Social | M, S+ |
| erkennen | .112 | Investigative | DI, I, S- |
| durchgreifen | .105 | Enterprising | M, DA, S+ |

Unter den zehn Verben, welche am stärksten positiv auf Faktor 4 luden, liessen sich Verben aus jeder RIASEC-Dimension mindestens einmal wiederfinden. Dabei war keine der Dimensionen mit mehr als drei Verben vertreten. Damit wurde deutlich, dass Faktor 4 keine einzelne Interessensdimension des Modells eindeutig widerspiegelte.

Am häufigsten vertreten waren Verben aus der Dimension Enterprising ($n = 3$), gefolgt von Verben aus den Dimensionen Investigative ($n = 2$) und Conventional ($n = 2$). Die stärkste positive Faktorladung von .578 wies dabei das Verb *führen* aus der Dimension

Enterprising auf. Die weiteren positiven Faktorladungen lagen deutlich unter diesem Wert und begannen ab einer Ladung von .291.

Die Verben verteilten sich gleichmäßig auf die Dimensionen der Dinge-Mensch-Achse (5 zu 5), die stärkste Faktorladung trug dabei jedoch ein Verb aus einer menschenorientierten Dimension. Entlang der Daten-Ideen-Achse überwogen die Verben aus den datenorientierten Dimensionen in ihrer Anzahl (5 zu 3) und auch in Hinsicht auf die stärkste positive Faktorladung.

Auf der Sociability-Achse überwogen die Verben aus den Dimensionen hoher Soziabilität sowohl in ihrer Anzahl (4 zu 3) als auch in Bezug auf die stärkste positive Faktorladung. Entlang der Conformity-Achse überwogen die Verben aus der Dimension hoher Konformität (2 zu 1). Die stärkste positive Faktorladung trug ein Verb, welches aus einer Dimension stammt, die nicht auf der Conformity-Achse liegt.

Tabelle 14

Bottom-10-Verben der 6-Faktorenlösung für Faktor 4.

| Verb | Faktorladung | RIASEC-Dimension | Achsen-Zuordnung |
|---------------|---------------------|-------------------------|-------------------------|
| leiten | -.227 | Enterprising | M, DA, S+ |
| reparieren | -.210 | Realistic | DI, S- |
| überreden | -.210 | Enterprising | M, DA, S+ |
| kontrollieren | -.195 | Conventional | DI, DA, C+ |
| ausstellen | -.150 | Artistic | M, I, C- |
| handeln | -.148 | Enterprising | M, DA, S+ |
| durchsetzen | -.147 | Enterprising | M, DA, S+ |
| trösten | -.136 | Social | M, S+ |
| zählen | -.131 | Conventional | DI, DA, C+ |
| reimen | -.123 | Artistic | M, I, C- |

Unter den zehn Verben, welche am stärksten negativ auf Faktor 4 luden, war mindestens ein Verb aus jeder RIASEC-Dimension, ausser aus Investigative, zu finden. Keine der Dimensionen trat dabei häufiger als viermal auf. Dies verdeutlichte, dass Faktor 4 sich auch nicht klar von einer einzelnen Interessensdimension des Modells abgrenzte.

Am häufigsten vertreten waren Verben der Dimension Enterprising ($n = 4$), gefolgt von Verben aus den Dimensionen Conventional ($n = 2$) und Artistic ($n = 2$). Hinsichtlich der Faktorladungen zeigte keines der Verben eine auffällig starke Negativladung im Vergleich zu den anderen Verben.

Auf der Dinge-Menschen-Achse überwogen die Verben aus den menschenorientierten Dimensionen (7 zu 3) und entlang der Daten-Ideen-Achse überwogen die Verben aus den datenorientierten Dimensionen in ihrer Anzahl (6 zu 2).

Die Verben aus den Dimensionen hoher Soziabilität, auf der Sociability-Achse, überwogen in ihrer Anzahl (5 zu 1) und auf der Conformity-Achse verteilten sich die Verben gleichmässig auf die Dimensionen hoher und niedriger Konformität (2 zu 2).

4.3.2.5 Faktor 5

In diesem Kapitel werden die Top-10 und Bottom-10 Faktorladungen für Faktor 5 präsentiert und beschrieben. Die zehn stärksten positiven Faktorladungen sind in Tabelle 15 und die zehn stärksten negativen in Tabelle 16 dargestellt.

Tabelle 15

Top-10-Verben der 6-Faktorenlösung für Faktor 5.

| Verb | Faktorladung | RIASEC-Dimension | Achsen-Zuordnung |
|-------------|---------------------|-------------------------|-------------------------|
| probieren | .267 | Investigative | DI, I, S- |
| malen | .237 | Artistic | M, I, C- |
| entdecken | .226 | Investigative | DI, I, S- |
| begleiten | .209 | Social | M, S+ |
| überzeugen | .198 | Enterprising | M, DA, S+ |
| durchsetzen | .175 | Enterprising | M, DA, S+ |
| ausdenken | .162 | Investigative | DI, I, S- |
| leiten | .156 | Enterprising | M, DA, S+ |
| vermuten | .142 | Investigative | DI, I, S- |
| auffordern | .129 | Enterprising | M, DA, S+ |

Unter den zehn Verben, welche am stärksten positiv auf Faktor 5 luden, liessen sich Verben aus jeder RIASEC-Dimension, ausser aus Realistic, mindestens einmal wiederfinden. Dabei war keine der Dimensionen mit mehr als vier Verben vertreten. Daran liess sich erkennen, dass Faktor 5 keine einzelne Interessensdimension des Modells eindeutig widerspiegelte.

Am häufigsten vertreten waren Verben aus den Dimensionen Investigative ($n = 4$) und Enterprising ($n = 4$). Hinsichtlich der Faktorladungen zeigt keines der Verben eine auffällig starke Positivladung im Vergleich zu den anderen Verben.

Die Verben aus den menschenorientierten Dimensionen auf der Dinge-Menschen-Achse (6 zu 4) und die Verben aus den ideenorientierten Dimensionen auf der Daten-Ideen-Achse (5 zu 4) überwogen in ihrer Anzahl.

Entlang der Sociability-Achse überwogen Verben aus den Dimensionen hoher Soziabilität in ihrer Anzahl (5 zu 4). Nur ein Verb entfiel auf eine Dimension der Conformity-Achse. Es handelte sich dabei um ein Verb aus der Dimension niedriger Konformität.

Tabelle 16

Bottom-10-Verben der 6-Faktorenlösung für Faktor 5.

| Verb | Faktorladung | RIASEC-Dimension | Achsen-Zuordnung |
|-------------|---------------------|-------------------------|-------------------------|
| beraten | -.308 | Social | M, S+ |
| aufbauen | -.275 | Realistic | DI, S- |
| beobachten | -.273 | Investigative | DI, I, S- |
| singen | -.271 | Artistic | M, I, C- |
| beantragen | -.214 | Conventional | DI, DA, C+ |
| verhandeln | -.169 | Enterprising | M, DA, S+ |
| gestalten | -.144 | Artistic | M, I, C- |
| umgraben | -.132 | Realistic | DI, S- |
| erkämpfen | -.125 | Enterprising | M, DA, S+ |
| pflanzen | -.103 | Realistic | DI, S- |

Unter den zehn Verben, welche am stärksten negativ auf Faktor 5 luden, befand sich mindestens ein Verb aus jeder RIASEC-Dimension. Keine der Dimensionen trat dabei

häufiger als dreimal auf. Damit wurde deutlich, dass Faktor 5 sich auch nicht klar von einer einzelnen Interessensdimension des Modells abgrenzte.

Am häufigsten vertreten waren Verben aus der Dimension Realistic ($n = 3$), gefolgt von Verben aus den Dimensionen Artistic ($n = 2$) und Enterprising ($n = 2$). Hinsichtlich der Faktorladungen zeigte keines der Verben eine auffällig starke Negativladung im Vergleich zu den anderen Verben.

Die Verben verteilten sich gleichmässig auf die Dimensionen der Dinge-Menschen-Achse (5 zu 5) und auch auf die Dimensionen der Daten-Ideen-Achse (3 zu 3).

Auf der Sociability-Achse überwogen die Verben aus den Dimensionen niedriger Soziabilität (4 zu 3) und entlang der Conformity-Achse überwogen die Verben aus der Dimension niedriger Konformität in ihrer Anzahl (2 zu 1).

4.3.2.6 Faktor 6

In diesem Kapitel werden die Top-10 und Bottom-10 Faktorladungen für Faktor 6 präsentiert und beschrieben. Die zehn stärksten positiven Faktorladungen sind in Tabelle 17 und die zehn stärksten negativen in Tabelle 18 dargestellt.

Tabelle 17

Top-10-Verben der 6-Faktorenlösung für Faktor 6.

| Verb | Faktorladung | RIASEC-Dimension | Achsen-Zuordnung |
|---------------|---------------------|-------------------------|-------------------------|
| erkennen | .559 | Investigative | DI, I, S- |
| probieren | .245 | Investigative | DI, I, S- |
| lehren | .243 | Social | M, S+ |
| gestalten | .221 | Artistic | M, I, C- |
| durchsetzen | .163 | Enterprising | M, DA, S+ |
| trösten | .156 | Social | M, S+ |
| beeinflussen | .154 | Enterprising | M, DA, S+ |
| ausdenken | .138 | Investigative | DI, I, S- |
| kontrollieren | .132 | Conventional | DI, DA, C+ |
| bohren | .131 | Realistic | DI, S- |

Unter den zehn Verben, welche am stärksten auf Faktor 6 luden, war mindestens ein Verb aus jeder RIASEC-Dimension zu finden. Dabei war keine Dimension mit mehr als drei Verben vertreten. Dies verdeutlichte, dass Faktor 6 keine einzelne Interessensdimension des Modells eindeutig widerspiegelte.

Am häufigsten vertreten waren Verben aus der Dimension Investigative ($n = 3$), gefolgt von Verben aus den Dimensionen Social ($n = 2$) und Enterprising ($n = 2$). Die stärkste positive Faktorladung von .559 zeigte dabei das Verb *erkennen* aus der Dimension Investigative. Die weiteren positiven Faktorladungen lagen deutlich unter diesem Wert und begannen ab einer Ladung von .245.

Die Verben verteilten sich gleichmässig auf die Dimensionen der Dinge–Menschen-Achse (5 zu 5), die stärkste positive Faktorladung trug dabei jedoch ein Verb aus einer dingeorientierten Dimension. Entlang der Daten-Ideen-Achse überwogen die Verben aus ideenorientierten Dimensionen in ihrer Anzahl (4 zu 3) und auch in Hinsicht auf die stärkste Faktorladung.

Auf der Sociability-Achse verteilten sich die Verben gleichmässig auf die Dimensionen hoher und niedriger Soziabilität (4 zu 4), die stärkste positive Faktorladung wies dabei aber ein Verb aus einer Dimension niedriger Soziabilität auf. Auch entlang der Conformity-Achse verteilten sich die Verben gleichmässig auf die Dimensionen hoher und niedriger Konformität (1 zu 1). Die stärkste positive Faktorladung trug ein Verb, welches aus einer Dimension stammt, die nicht auf der Conformity-Achse liegt.

Tabelle 18

Bottom-10-Verben der 6-Faktorenlösung für Faktor 6.

| Verb | Faktorladung | RIASEC-Dimension | Achsen-Zuordnung |
|-----------------|---------------------|-------------------------|-------------------------|
| beschützen | -.247 | Social | M, S+ |
| begleiten | -.245 | Social | M, S+ |
| überzeugen | -.231 | Enterprising | M, DA, S+ |
| beobachten | -.162 | Investigative | DI, I, S- |
| malen | -.136 | Artistic | M, I, C- |
| hämmern | -.123 | Realistic | DI, S- |
| rätseln | -.119 | Investigative | DI, I, S- |
| auskundschaften | -.108 | Investigative | DI, I, S- |
| ordnen | -.101 | Conventional | DI, DA, C+ |
| umgraben | -.097 | Realistic | DI, S- |

Auch unter den zehn Verben, welche am stärksten negativ auf Faktor 6 luden, waren Verben aus allen RIASEC-Dimensionen mindestens einmal vertreten. Keine der Dimensionen trat dabei häufiger als dreimal auf. Daran liess sich erkennen, dass sich Faktor 6 auch nicht klar von einer einzelnen Interessensdimension des Modells abgrenzte.

Am häufigsten vertreten waren Verben aus der Dimension Investigative ($n = 3$), gefolgt von Verben aus den Dimensionen Social ($n = 2$) und Realistic ($n = 2$). Hinsichtlich der Faktorladungen zeigte keines der Verben eine auffällig starke Negativladung im Vergleich zu den anderen Verben.

Auf der Dinge-Menschen-Achse überwogen die Verben aus den dingeorientierten Dimensionen (6 zu 4) und entlang der Daten-Ideen-Achse überwogen die Verben aus den ideenorientierten Dimensionen in ihrer Anzahl (4 zu 2).

Auf der Sociability-Achse überwogen die Verben aus den Dimensionen niedriger Soziabilität in ihrer Anzahl (5 zu 3). Entlang der Conformity-Achse verteilten sich die Verben gleichmässig auf die Dimensionen hoher und niedriger Konformität (1 zu 1).

5 Interpretation und Diskussion

In diesem Kapitel werden die Ergebnisse dieser Arbeit interpretiert und diskutiert. Das Kapitel ist aufgeteilt in eine Interpretation der einzelnen extrahierten Faktoren in Bezug auf das RIASEC-Modell, einer Diskussion der Ergebnisse, den Limitationen der Arbeit und einem Forschungsausblick.

5.1 Interpretation der Faktoren

Die Interpretationen der einzelnen Faktoren erfolgt anhand der Dimensionen des RIASEC-Modells. Für jeden Faktor wird zunächst beschrieben, ob dieser eine der RIASEC-Dimensionen widerspiegelt oder welcher er am ehesten entspricht. Danach erfolgt eine vertiefte inhaltliche Interpretation des Faktors anhand der Mittelwerte und den Ergebnissen der zehn stärksten positiven und negativen Faktorladungen.

5.1.1 Faktor 1

Die Ergebnisse aus den Mittelwerten der Faktorladungen und den zehn stärksten positiven sowie negativen Faktorladungen verdeutlichen, dass Faktor 1 keine der Interessensdimensionen des RIASEC-Modells eindeutig widerspiegelt. Die grösste Übereinstimmung zeigt der Faktor aber mit der Dimension Investigative. Dies aufgrund des höchsten Mittelwerts dieser Dimension, verglichen mit den anderen Faktoren als auch innerhalb des Faktors. Hinzu kommt, dass die stärkste positive Faktorladung ein Verb aus dieser Dimension aufweist.

Zur Beschreibung der inhaltlichen Struktur von Faktor 1, bietet die Betrachtung der Mittelwerte der RIASEC-Dimensionen erste Hinweise. Im Gesamtvergleich der sechs extrahierten Faktoren zeigt Faktor 1 die höchsten Mittelwerte für die Dimensionen Investigative und Realistic und den niedrigsten Mittelwert für die Dimension Social. Der innerhalb des Faktors höchste Mittelwert liegt in der Dimension Investigative und der

niedrigste Mittelwert zeigt die Dimension Social. Die Mittelwerte weisen darauf hin, dass sich der inhaltliche Charakter dieses Faktors, von den sozialen unterstützenden Tätigkeiten entfernt und stärker an systematischen, beobachtenden und praktischen Tätigkeiten orientiert.

Weiter gestützt wird diese Tendenz durch die Häufigkeit der in den zehn stärksten positiven Faktorladungen vertretenen RIASEC-Dimensionen. Diese werden dominiert durch Verben aus den Dimensionen Realistic, Investigative und Conventional. Die Häufigkeit der Verben aus der Dimension Conventional unterstreicht weiter die Tendenz für systematische Tätigkeiten. In den zehn stärksten negativen Faktorladungen zeigt sich eine übermäßige Verteilung in den gegensätzlichen Dimensionen. Insbesondere Verben aus der Dimension Social finden sich häufig wieder. Dies bestätigt weiter die Abgrenzung des Faktors von sozialen und unterstützenden Tätigkeiten.

Zusätzliche Hinweise zur semantischen Zusammensetzung des Faktors ergeben sich aus der Betrachtung der Achsen, welche dem RIASEC-Modell zugrunde liegen. Auf der Dinge-Mensch-Achse zeigt sich eine Dominanz zu dingeorientierten und eine Abgrenzung von menschenorientierten Dimensionen. Auch die Verteilung entlang der Sociability-Achse zeigt eine Tendenz entgegen Dimensionen hoher Soziabilität und hin zu Dimensionen niedriger Soziabilität. Diese Ergebnisse legen nahe, dass der Faktor Tätigkeiten beschreibt, welche mit minimaler oder gar keiner sozialen Interaktion verknüpft sind und hauptsächlich den Umgang mit Dingen oder Gegenständen erfordern. Auf der Daten-Ideen-Achse deutet die Verteilung auf eine leichte Tendenz zu ideenorientierten Tätigkeiten hin. Diese Tendenz erklärt sich aber insbesondere durch das stärkste positiv ladende Verb *finden*, welches aus der ideenorientierten Dimension Investigative stammt. Die Verteilung auf der Conformity-Achse ist beinahe ausgeglichen, mit einer leichten Tendenz hin zu Tätigkeiten hoher Konformität.

Der semantische Kern des Faktors lässt sich über die deutlich stärkste positive Faktorladung des Verbs *finden* aus der Dimension Investigative beschreiben. Es repräsentiert eine explorative und kognitive Auseinandersetzung mit Inhalten. Zusammen mit den positiv

ladenden Verben *folgern*, *einordnen* und *beantragen* zeichnet der Faktor sich weiter durch systematische, strukturierte und sachlich orientierte Tätigkeiten aus. Weiter ergänzt wird dieses Bild durch positive Ladungen der Verben *pflanzen*, *schrauben* und *reparieren*, welche ebenfalls klar definierte, gegenstandsbezogene und zusätzlich praktische Tätigkeiten umfassen. Die negativ ladenden Verben *lehren*, *beraten* und *empfinden* stammen aus Dimensionen, welche von der Bedeutung her eher den Gegensatz zu den eben beschriebenen Tätigkeiten bilden. Etwas auffällig in den positiven Ladungen ist das Verb *begleiten* aus der Dimension Social, welches die viertstärkste positive Faktorladung aufweist. Dies könnte dadurch erklärt werden, dass dieses Verb nicht nur in sozialen Kontexten verwendet wird.

Für Faktor 1 ergibt sich insgesamt ein semantischer Raum, der systematische, strukturierte und praktische Tätigkeiten, die eine explorative Auseinandersetzung mit Dingen erfordern, beschreibt. Der Faktor grenzt sich dabei inhaltlich klar von menschenorientierten, sozialen und unstrukturierten Tätigkeiten ab.

5.1.2 Faktor 2

Die Betrachtung der Mittelwerte sowie der zehn stärksten positiven und negativen Faktorladungen zeigt, dass Faktor 2 keine der RIASEC-Dimensionen eindeutig widerspiegelt. Dabei weist der Faktor jedoch die stärkste Ähnlichkeit mit der Dimension Social auf. Dies zeigt sich im höchsten Mittelwert dieser Dimension, sowohl innerhalb des Faktors als auch im Vergleich mit den anderen Faktoren. Zudem stammt die stärkste positive Faktorladung ebenfalls von einem Verb aus dieser Dimension.

Erste wesentliche Merkmale zur inhaltlichen Bedeutung von Faktor 2 ergeben sich aus der Betrachtung der Mittelwerte der RIASEC-Dimensionen. Darunter weist die Dimension Social nicht nur innerhalb des Faktors sondern auch im Gesamtvergleich mit den sechs extrahierten Faktoren den höchsten Mittelwert auf. Der niedrigste Mittelwert innerhalb des Faktors zeigt sich für die Dimension Enterprising. Auffällig ist, dass der Faktor ausserdem

zusätzlich für die Dimensionen Investigative und Conventional den niedrigsten Mittelwert über alle Faktoren besitzt. Diese Ergebnisse legen nahe, dass der semantische Charakter des Faktors sich deutlich auf zwischenmenschliche Tätigkeiten, insbesondere im Sinn von Unterstützung, fokussiert. Gleichzeitig grenzt sich der Faktor aber von zwischenmenschlichen Tätigkeiten im unternehmerischen Sinn ab. Ebenfalls distanziert sich der Faktor von systematischen, beobachtenden und strukturierten Tätigkeiten.

Auch die Verteilung der Verben innerhalb der zehn stärksten positiven und negativen Faktorladungen bestärkt diese Interpretation. In den positiven Ladungen zeigt sich eine Tendenz zu menschenorientierten und unterstützenden Tätigkeiten, da Verben aus der Dimension Social am häufigsten vertreten sind. Die Häufigkeit der Verben aus den Dimensionen Enterprising und Realistic sind in den positiven sowie negativen Ladungen ausgeglichen und relativieren dadurch den Einfluss ihrer Verteilung auf den Inhalt des Faktors. Unter den negativen Ladungen stammen jedoch auffällig viele Verben aus der Dimension Investigative was für eine Abgrenzung von explorativen, symbolischen und beobachtenden Tätigkeiten spricht.

Entlang der dem RIASEC-Modell zugrundeliegenden Achsen lassen sich weitere diese Annahme stützende Hinweise finden. Auf der Dinge-Menschen-Achse zeigt sich eine Verteilung in die Richtung von menschenorientierten und entgegen dingeorientierter Dimensionen. Faktor 2 zeigt aufgrund dessen eine deutliche Abgrenzung zu den gegenstandsbezogenen Tätigkeiten und eine Tendenz zu Tätigkeiten, welche eine zwischenmenschliche Komponente mit sich bringen. Die Verteilung der Faktorladungen auf der Sociability-Achse bestätigt diese Tendenzen. In Faktor 2 verteilt sich nur eine sehr geringe Anzahl der Verben auf die Pole der Conformity-Achse und weist eine leichte Tendenz in Richtung von Tätigkeiten von hoher Konformität auf.

Der semantische Kern des Faktors bildet das Verb *helfen* aus der Dimension Social mit der stärksten positiven Faktorladung. Gemeinsam mit den positiv ladenden Verben

beraten und *begleiten* repräsentieren sie zwischenmenschliche, kommunikative Tätigkeiten zur Unterstützung. Im Gegensatz dazu grenzt sich der Faktor von klar definierten, strukturierten und beeinflussenden Tätigkeiten ab. Dies verdeutlicht sich durch die negativ ladenden Verben *überzeugen*, *handeln*, *erkennen* und *einordnen*. Etwas auffällig ist die zweitstärkste positive Faktorladung des Verbs *testen* und die viertstärkste positive Faktorladung des Verbs *entdecken* aus der Dimension Investigative. Diese Verben scheinen jedoch sehr abstrakt und in unterschiedlichen Kontexten anwendbar.

Insgesamt ergibt sich für Faktor 2 somit ein semantisches Profil, welches sich durch zwischenmenschliche Tätigkeiten, kommunikative Interaktion und Unterstützung auszeichnet. Er grenzt sich aber klar von Tätigkeiten ab, welche eine starke Strukturierung und Systematik voraussetzen oder von unternehmerischer und beeinflussender Natur sind.

5.1.3 Faktor 3

Für Faktor 3 zeigen die Ergebnisse aus den Mittelwerten und den zehn stärksten positiven und negativen Faktorladungen, dass keine der Interessensdimensionen des RIASEC-Modells eindeutig widerspiegelt wird. Die grösste Passung zeigt der Faktor dabei mit der Dimension Artistic. Dies lässt sich durch die Häufigkeit von Verben aus dieser Dimension in den stärksten positiven und negativen Faktorladungen erklären. Der Faktor weist darüber hinaus für diese Dimension auch den höchsten Mittelwert auf, sowohl innerhalb des Faktors als auch im Vergleich mit den anderen Faktoren.

Zur Beschreibung der inhaltlichen Struktur von Faktor 3 ergeben sich erste Hinweise aus der Betrachtung der Mittelwerte der RIASEC-Dimensionen. Innerhalb des Faktors weist die Dimension Artistic den höchsten Mittelwert auf, welches auch der höchste Mittelwert dieser Dimension im Vergleich zu den anderen Faktoren ist. Auffällig ist, dass dieser Faktor verglichen mit den anderen Faktoren ebenfalls den höchsten Mittelwert für die Dimension Conventional aufweist. Die beiden Dimensionen gelten im RIASEC-Modell als

gegensätzlich. Anhand dieser Mittelwerte repräsentiert der Faktor unstrukturierte und freie als auch stark strukturierte und systematische Tätigkeiten. Der innerhalb des Faktors niedrigste Mittelwert weist die Dimension Investigative auf. Aus der Betrachtung der Mittelwerte lässt sich keine eindeutige Interessensstruktur erkennen. Jedoch weist der deutlich höhere Mittelwert der Dimension Artistic auf eine stärkere Tendenz in die Richtung von kreativen und künstlerischer Tätigkeiten hin. Gleichzeitig grenzt sich der Faktor durch den niedrigsten Mittelwert von explorativen und beobachtenden Tätigkeiten ab.

In den zehn stärksten positiven Faktorladungen sind am häufigsten Verben aus den Dimensionen Artistic und Enterprising vertreten, während nur ein einziges Verb aus der Dimension Conventional stammt. In den zehn stärksten negativen Faktorladungen ist die Dimension Conventional jedoch gemeinsam mit der Dimension Enterprising am häufigsten vertreten. Diese Verteilung betont die Annahme, dass der Faktor eine stärkere Tendenz zu künstlerischen und unstrukturierten als zu strukturierten und systematischen Tätigkeiten zeigt. Da unter den negativen Faktorladungen die Dimension Enterprising gleich oft wie in den positiven Faktorladungen vertreten ist, lässt sich aus der reinen Verteilung kein klarer Schluss ziehen.

Die Betrachtung der zugrundeliegenden Achsen des RIASEC-Modells unterstreicht die Mehrdeutigkeit des Faktors. Auf der Dinge-Mensch-Achse zeigt sich insgesamt eine leichte Tendenz in die Richtung von menschenorientierten Dimensionen. Entlang der Sociability-Achse zeigt die Verteilung jedoch, dass der Faktor sich von zwischenmenschlichen Tätigkeiten eher abgrenzt. Die Verteilung auf der Daten-Ideen-Achse zeigt eine Abgrenzung von datenorientierten Dimensionen. Dies weist darauf hin, dass sich der inhaltliche Charakter des Faktors nicht durch datenorientierte Tätigkeiten beschreiben lässt, was sich mit der Annahme zur kreativen und künstlerischen Tendenz des Faktors deckt. Die Verteilung auf der Conformity-Achse ist ebenfalls passend zu dieser Tendenz. Dies zeigt

sich durch eine Neigung hin zu Tätigkeiten niedriger Konformität bei gleichzeitiger Abgrenzung von Tätigkeiten hoher Konformität.

Der semantische Kern des Faktors lässt sich nicht klar anhand der stärksten positiven Faktorladung des Verbs *zählen* aus der Dimension Conventional beschreiben. Da wie zuvor erwähnt, diese Dimension als Gegensatz zur Dimension Artistic gilt und somit den bisherigen Annahmen widersprechen würde. Die Zusammensetzung des Faktors aus den positiv ladenden Verben *gestalten, malen, singen*, sowie *handeln, investieren* und *beeinflussen* deutet auf eine Verbindung von kreativem Ausdruck und eigenverantwortlichem unternehmerischen Handeln hin. Dabei weist die stärkste Faktorladung auf eine zusätzliche Komponente der Datenverarbeitung hin. Im Gegensatz dazu grenzt sich der Faktor inhaltlich durch die negativ ladenden Verben *führen, durchsetzen* und *betreuen* von Tätigkeiten der Führung oder Beeinflussung anderer Menschen ab. Hinzu kommt eine Abgrenzung von Strukturierung und Systematik durch die negativ ladenden Verben *ordnen, aussortieren* und *beantragen*. Eine Auffälligkeit zeigt die viertstärkste positive Faktorladung des Verbs *entdecken* aus der Dimension Investigative. Wie bereits in der Interpretation zu Faktor 2 erwähnt, scheint dies ein eher abstraktes Verb zu sein, welches in unterschiedlichen Kontexten Anwendung findet.

Insgesamt ergibt sich für Faktor 3 so ein semantischer Raum, der Tätigkeiten des kreativen Ausdrucks, selbstbestimmtem unternehmerischen Handelns und die Verarbeitung von Daten abbildet. Währenddessen grenzt sich der Faktor inhaltlich aber von strukturierenden und systematischen Tätigkeiten, sowie der Führung oder Beeinflussung von Menschen ab.

5.1.4 Faktor 4

Aus den Ergebnissen der Mittelwerte der Faktorladungen und den zehn stärksten positiven und negativen Faktorladungen wurde deutlich, dass Faktor 4 keine der Interessensdimensionen des RIASEC-Modells eindeutig widerspiegelt. Der Faktor scheint mit

keiner der Dimensionen eine besonders hohe Übereinstimmung zu zeigen Innerhalb des Faktors weist die Dimension Investigative den höchsten Mittelwert auf. Jedoch ist der Mittelwert dieser Dimension in drei anderen Faktoren deutlich höher als für Faktor 4.

Die Betrachtung der Mittelwerte der RIASEC-Dimensionen bietet erste Hinweise zur inhaltlichen Struktur von Faktor 4. Innerhalb des Faktors weist die Dimension Investigative den höchsten und die Dimension Artistic den niedrigsten Mittelwert auf. Anhand der Mittelwerte innerhalb des Faktors zeigt sich eine inhaltliche Tendenz zu explorativen und beobachtenden Tätigkeiten. Dabei grenzt sich der Faktor ab von Tätigkeiten welche unstrukturiert und künstlerisch sind.

Weitere Hinweise zur inhaltlichen Struktur des Faktors ergeben sich aus der Verteilung der zehn stärksten positiven und negativen Faktorladungen. Die Dimension Enterprising ist jeweils in den positiv sowie den negativ ladenden Verben am häufigsten vertreten, wobei in den negativen Faktorladungen ein Verb mehr aus dieser Dimension stammt. Auch Verben aus der Dimension Conventional sind in den positiven und negativen Faktorladungen vertreten, kommen aber gleich häufig vor. Neben dieser zweideutigen Verteilungen finden sich in den positiven Ladungen noch Verben aus der Dimension Investigative, während Verben dieser Dimension in den negativen Ladungen nicht vorhanden sind. Währenddessen sind Verben aus der Dimension Artistic in den negativen Ladungen häufiger vertreten. Diese Verteilung unterstützt Annahme zur Tendenzen des Faktor zu explorativen und beobachtenden Tätigkeiten und einer Abgrenzung von unstrukturierten und künstlerischen Tätigkeiten.

Die Untersuchung der dem RIASEC-Modell zugrundeliegenden Achsen liefert weitere Hinweise zur inhaltlichen Deutung des Faktors. Die Verteilung auf der Dinge-Menschen-Achse und der Sociability-Achse legt nahe, dass der Faktor sich von menschenorientierten und zwischenmenschlichen Tätigkeiten abgrenzt. Entlang der Daten-Ideen-Achsen zeigt sich kein klares Ergebnis, da unter den positiven und den negativen Ladungen jeweils die

Mehrheit der Verben aus datenorientierten Dimensionen stammen. Die Conformity-Achse zeigt ein ähnlich unklares Bild, weist jedoch eine leichte Tendenz zu Tätigkeiten hoher Konformität auf. Diese Tendenz passt zur inhaltlichen Abgrenzung des Faktors von unstrukturierten Tätigkeiten.

Die stärkste Faktorladung stammt vom Verb *führen* aus der Dimension Enterprising. Dies erschwert die Beschreibung des semantischen Kerns des Faktors, da die Dimension Enterprising im RIASEC-Modell den Gegensatz zur Dimension Investigative bildet und somit den bisherigen Annahmen widerspricht. Die Zusammensetzung des Faktors aus den positiv ladenden Verben *entdecken*, *verwalten*, *kalkulieren* und *erkennen* deuten auf eine Repräsentation von systematischen, explorativen und strukturierten Tätigkeiten hin. Während die negativ ladenden Verben *kontrollieren* und *zählen* gleichzeitig auf eine Abgrenzung von überprüfenden Tätigkeiten hindeuten. In den positiven Faktorladungen finden sich zudem die Verben *führen*, *fordern* und *durchgreifen* und in den negativen Faktorladungen die Verben *leiten*, *überreden*, *handeln* und *durchsetzen* aus der Dimension Enterprising. Diese Verben weisen daraufhin, dass der Faktor innerhalb von Tätigkeiten der Einflussnahme differenziert. Die positiv ladenden Verben beziehen sich dabei eher auf Tätigkeiten, bei welchen bereits Macht oder Autorität vorhanden ist und gezielt zur Einflussnahme genutzt wird. Die negativ ladenden Verben zeichnen sich dabei eher durch eine Einflussnahme basierend auf Strategie und Überzeugung aus.

Für Faktor 4 ergibt sich insgesamt ein semantischer Charakter der sich auf explorative, strukturierte, systematische und beobachtende Tätigkeiten fokussiert. Ergänzt wird dieser durch eine Komponente direkter Einflussnahme, die auf Macht und Autorität basiert. Eine inhaltliche Abgrenzung zeigt sich gegenüber unstrukturierten, künstlerischen und überprüfenden sowie menschenorientierten oder zwischenmenschlichen Tätigkeiten.

5.1.5 Faktor 5

Die Ergebnisse aus den Mittelwerten der Faktorenladungen und den zehn stärksten positiven sowie negativen Faktorladungen verdeutlichen, dass Faktor 5 keine der Interessensdimensionen des RIASEC-Modells eindeutig widerspiegelt. Die grösste Ähnlichkeit zeigt sich mit der Dimension Investigative. Dies ergibt sich aus der Häufigkeit von Verben dieser Dimension in den zehn stärksten positiven Faktorladungen und dem höchsten Mittelwert der Dimension innerhalb des Faktors.

Zur Beschreibung der inhaltlichen Struktur von Faktor 5, finden sich erste Hinweise in den Mittelwerten der RIASEC-Dimensionen. Innerhalb des Faktors weist die Dimension Investigative den höchsten und die Dimension Realistic den niedrigsten Mittelwert auf. Im Vergleich mit allen sechs Faktoren zeigt Faktor 5 für die Dimensionen Realistic und Artistic den niedrigsten Mittelwert. Die Mittelwerte deuten daraufhin, dass der Faktor zu explorativen, beobachtenden und systematischen Tätigkeiten tendiert. Währenddessen grenzt sich der Faktor von praktischen, künstlerischen und unstrukturierten Tätigkeiten ab.

Die Häufigkeiten innerhalb der zehn stärksten positiven und negativen Faktorladungen unterstützen und erweitern diese Tendenz. Die positiven Faktorladungen setzen sich beinahe ausschliesslich und zu gleichen Teilen aus Verben der Dimensionen Investigative und Enterprising zusammen. Dies ist insofern auffällig, da die beiden Dimensionen im RIASEC-Modell Gegensätze darstellen. Die Häufigkeit der Verben aus der Dimension Enterprising ergänzt somit die Annahme zur inhaltlichen Struktur des Faktors um Tätigkeiten sozialer Beeinflussung im unternehmerischen Kontext. In den zehn stärksten negativen Faktorladungen unterstreichen die Häufigkeiten der Verben aus den Dimensionen Realistic und Artistic die inhaltliche Abgrenzung des Faktors von praktischen, künstlerischen und unstrukturierten Tätigkeiten. Weiter finden sich auch Verben aus der Dimension Enterprising wieder, was auf eine Differenzierung innerhalb von Tätigkeiten der sozialen Beeinflussung im unternehmerischen Kontext hinweist.

Entlang der dem RIASEC-Modell zugrundeliegenden Achsen ergeben sich weitere Hinweise zur inhaltlichen Interpretation des Faktors. Die Verteilung auf der Dinge-Menschen-Achse und der Sociability-Achse zeigen eine leichte Tendenz zu menschenorientierten und zwischenmenschliche Tätigkeiten. Auf der Daten-Ideen zeigt sich ebenfalls nur eine leichte Tendenz, in die Richtung von ideenorientierten Tätigkeiten. Die Verteilung entlang der Conformity-Achse ist durchmischt und zeigt sowohl eine leichte Tendenz zu als auch eine leichte Abgrenzung von Tätigkeiten niedriger Konformität. Diese Verteilung auf den zugrundeliegenden Achsen entspricht grösstenteils den bisherigen Annahmen zu Faktor 5.

Die Stärke der zehn positivsten und zehn negativsten Faktorladungen dieses Faktors beinhalten einige Auffälligkeiten. Einerseits liegt die stärkste positive Faktorladung, des Verbs *probieren* aus der Dimension Investigative nur leicht über der zweitstärksten Faktorladung des Verbs *malen* aus der Dimension Artistic. Andererseits ist Faktor 5 der einzige Faktor, bei welchem mehrere negative Ladungen betragsmässig über der stärksten positiven Ladung liegen. Anhand dieser Ergebnisse lässt sich der semantische Kern des Faktors eher durch die negativ ladenden Verben beschreiben. Die stärkste negative Faktorladung stammt vom Verb *beraten* aus der Dimension Social. Dadurch grenzt sich der Faktor inhaltlich von kommunikativ unterstützenden Tätigkeiten ab. Die weiteren negativen Faktorladungen der Verben *aufbauen*, *umgraben*, *pflanzen*, *singen* und *gestalten* stützen die anfangs getroffenen Annahmen zur Abgrenzung von körperlichen und künstlerischen Tätigkeiten. Die positiv ladenden Verben *probieren*, *entdecken*, *ausdenken* und *vermuten* zeigen eine inhaltliche Tendenz des Faktors zu explorativen und experimentellen Tätigkeiten. Diese Tendenz ergänzt sich um ein Element der Einflussnahme durch die weiteren positiv ladenden Verben *überzeugen*, *durchsetzen*, *auffordern* und *leiten*. Während die negativ ladenden Verben *verhandeln* und *erkämpfen* auf eine Differenzierung von Tätigkeiten hinweisen, bei denen Einfluss gewonnen werden muss.

Insgesamt ergibt sich für Faktor 5 ein semantisches Profil welches sich von praktischen, körperlichen, künstlerischen und kommunikativ unterstützenden Tätigkeiten abgrenzt. Dabei beschreibt der Faktor Tätigkeiten, bei welchen die Einflussnahme mit explorativen und experimentellen Tätigkeiten kombiniert wird.

5.1.6 Faktor 6

Die Betrachtung der Mittelwerte sowie der zehn stärksten positiven und negativen Faktorladungen zeigt, dass Faktor 6 keine der RIASEC-Dimensionen eindeutig widerspiegelt. Dabei zeigt der Faktor die höchste Übereinstimmung mit der Dimension Investigative. Dies lässt sich durch die stärkste positive Faktorladung und den höchsten Mittelwert dieser Dimension innerhalb des Faktors erklären. Der Mittelwert liegt dabei nur knapp unter dem Mittelwert des Faktors, welcher im Vergleich zu den anderen Faktor den höchsten Mittelwert für diese Dimension aufweist.

Erste Hinweise zur inhaltlichen Ausprägung von Faktor 6 ergeben sich aus der Betrachtung der Mittelwerte der RIASEC-Dimensionen. Innerhalb des Faktors weist die Dimension Investigative den höchsten und die Dimension Social den niedrigsten Mittelwert auf. Anhand dieser Mittelwerte tendiert der Faktor inhaltlich zu explorativen, systematischen und beobachtenden Tätigkeiten und grenzt sich von zwischenmenschlichen unterstützenden Tätigkeiten ab. Weiter zeigt die Dimension Enterprising für diesen Faktor im Vergleich zu allen sechs extrahierten Faktoren den höchsten Mittelwert. Dies weist darauf hin, dass der Faktor doch eine Tendenz zu zwischenmenschlichen Tätigkeiten zeigt, die sich jedoch auf den unternehmerischen Kontext beziehen.

Die Häufigkeiten innerhalb der zehn stärksten positiven und negativen Faktorladungen zeigen eher eine zweideutige Verteilung. In den positiven wie auch den negativen Faktorladungen stammen die meisten und jeweils die gleiche Anzahl Verben aus der Dimension Investigative, gefolgt von Verben aus der Dimension Social. Dies zeigt einen

Widerspruch in sich selbst und aus der reinen Verteilung lässt sich kein klarer Schluss ziehen. Es weist jedoch auf eine Differenzierung innerhalb der Tätigkeiten dieser Dimensionen hin. Die Häufigkeit der Verben aus der Dimension Enterprising innerhalb der positiven Faktorladungen stützt die Annahme der Tendenz zwischenmenschlichen Tätigkeiten im unternehmerischen Kontext. Eine weiterer Hinweis zum semantischen Charakter des Faktors zeigt die Häufigkeit der Verben aus der Dimension Realistic innerhalb der negativ ladenden Verben. Diese Verteilung deutet auf eine Abgrenzung von praktischen Tätigkeiten hin.

Die Verteilung auf der Dinge-Menschen-Achse und der Sociability-Achse weisen auf eine Abgrenzung von dingeorientierten Tätigkeiten hin. Entlang der Daten-Ideen-Achse entstehen Widersprüche, da die positiven Faktorladungen einen Tendenz zu ideenorientierten und die negativen Faktorladungen eine Abgrenzung von ideenorientierten Tätigkeiten zeigen. Auf der Conformity-Achse zeigt sich eine gleichmässige Verteilung, woraus sich weder eine Tendenz für noch eine Abgrenzung von Tätigkeiten hoher beziehungsweise niedriger Konformität ergibt.

Der semantische Kern des Faktors lässt sich anhand der deutlich stärksten positiven Faktorladung des Verbs *erkennen* aus der Dimension Investigative beschreiben. Gemeinsam mit den positiv ladenden Verben *probieren* und *ausdenken* beschreibt der Faktor Tätigkeiten mit einem explorativen, zielgerichteten und analytischen Charakter. Gleichzeitig differenziert der Faktor durch die negativ ladenden Verben *beobachten*, *rätseln* und *auskundschaften* von explorativen Tätigkeiten die sich als offen, unsystematisch und passiv beschreiben lassen. Die positiv ladenden Verben *lehren* und *trösten* stehen im Widerspruch zur Abgrenzung des Faktors von zwischenmenschlich unterstützenden Tätigkeiten. Insbesondere das Verb *trösten* lässt sich kaum in einen nicht-sozialen unterstützenden Kontext übertragen und bleibt ein bestehender Widerspruch zu den Annahmen des Faktors. Wohingegen das Verb *lehren* auch in einem unternehmerischen Kontext Anwendung finden könnte. Die negativ ladenden Verben *beschützen* und *begleiten* passen ebenfalls in den sozial unterstützenden Kontext, von

welchem sich der Faktor abgrenzt. Weiter ergänzt wird dieses Bild durch positiv ladende Verben *durchsetzen* und *beeinflussen*, die auf eine zusätzliche Komponente der Überzeugung hinweisen. Durch die negativ ladenden Verben *hämmern* und *umgraben* zeigt sich eine Abgrenzung zu praktischen Tätigkeiten.

Insgesamt ergibt sich für Faktor 6 ein semantisches Profil, welches sich durch eine Verbindung von explorativen zielgerichteten und analytischen Tätigkeiten mit der Komponente von zwischenmenschlicher Überzeugung auszeichnet. Weiter grenzt der Faktor sich eher von zwischenmenschlich unterstützenden und praktischen Tätigkeiten ab.

5.2 Diskussion

Aus der Interpretation der sechs extrahierten Faktoren haben sich unterschiedliche Ergebnisse zur Abbildung der Dimensionen des RIASEC-Modells ergeben. Eine eindeutige Übereinstimmung zwischen einzelnen Faktoren und einzelnen Dimensionen zeigt sich dabei in keinem Fall. In den Faktoren lassen sich jedoch Tendenzen und Überschneidungen zum Modell sowie Hinweise auf erweiterte und komplexere Interessensräume finden. Im Folgenden werden zentrale Auffälligkeiten aus der Analyse und Interpretation zusammenfassend diskutiert.

Ein erstes auffälliges Muster zeigt sich für die Dimension Investigative. In 4 von 6 Faktoren trägt diese Dimension den höchsten Mittelwert und bei drei Faktoren gehört die stärkste positive Faktorladung zu einem Verb aus dieser Dimension. Hinzu kommt, dass die Dimensionen Investigative und Enterprising besonders häufig in den zehn stärksten positiven und negativen Faktorladungen vertreten sind. Dabei sind die Verben der Dimension Investigative vor allem in den positiven Ladungen zu finden, während sich die Verben der Dimension Enterprising gleichmässiger auf positive und negative Ladungen verteilen. Verben aus diesen Dimensionen sind oftmals abstrakt, prozessbezogen und in unterschiedlichen Kontexten vorzufinden. Darüber hinaus könnte diese Überrepräsentation auch ein Hinweis

darauf sein, dass Verben dieser Dimensionen eine übergeordnete Bedeutung für unterschiedliche Berufsfelder besitzen. So könnten beispielsweise Tätigkeiten aus der Dimension Investigative in einer grossen Anzahl von sehr unterschiedlichen Berufen ergänzend zur Haupttätigkeit ausgeübt werden.

Die am deutlichsten interpretierbaren semantischen Strukturen weisen die Faktoren 1 und 2 auf. Sie teilen sich inhaltlich auf in Dimensionen, die sich entweder auf systematische, strukturierte und praktische oder zwischenmenschliche unterstützende Tätigkeiten konzentrieren. Gleichzeitig lässt sich jeweils eine Abgrenzung zu Tätigkeiten erkennen, welche diesen Tendenz nicht entsprechen. Diese beiden Faktoren stellen eine Art Polarisierung dar, welche der Gegenüberstellung der menschen- und dingeorientierten Dimensionen des RIASEC-Modells nach Prediger (1982) gleicht. In den Faktoren 3, 4, 5 und 6 treten eher semantische Mischformen und teilweise Überlagerungen von Bedeutungen auf. Häufig tauchen in einem Faktor Tendenzen aus gegensätzlichen Dimensionen des RIASEC-Modells auf. So kombiniert beispielsweise Faktor 3 die unstrukturierten, kreativen und künstlerischen Tätigkeiten der Dimension Artistic mit der gegenteiligen Dimension Conventional, die den systematischen und strukturierten Umgang mit Daten beschreibt. Faktor 4, 5 und 6 verbinden jeweils die gleichen gegensätzlichen Dimensionen des Modells. Alle diese Faktoren kombinieren eine Art von explorativen Tätigkeiten aus der Dimension Investigative, mit Tätigkeiten der Einflussnahme oder Überzeugung aus der Dimension Enterprising. Dabei unterscheiden sich diese drei Faktoren einerseits durch feinere Differenzierungen innerhalb dieser Tätigkeiten und andererseits durch Einflüsse aus anderen Dimensionen des RIASEC-Modells.

Diese Faktoren könnten Hinweise auf komplexere Interessensräume sein, welche durch das RIASEC-Modell nicht oder unzureichend dargestellt werden. Die Forschungsfrage inwiefern künstliche Intelligenz die menschliche Realität von Interessensstrukturmodellen

insbesondere dem RIASEC-Modell widerspiegelt, lässt sich demnach folgendermassen beantworten:

Die sechs Faktoren, welche auf der Basis des VITs aus dem Word-Embedding extrahiert wurden, spiegeln die Interessensdimensionen des RIASEC-Modells nicht wider. Die Faktoren stimmen mit keiner der Dimensionen eindeutig überein. Es lassen sich jedoch teilweise dem Modell ähnliche Tendenzen in den Faktoren wiederfinden. Die Hypothese der vorliegenden Arbeit, dass künstliche Intelligenz nur einen Teil der menschlichen Realität des RIASEC-Modells widerspiegelt, aber noch weitere eigene Dimensionen bildet konnte teilweise bestätigt werden. Dabei ist wichtig zu erkennen, dass keine komplett neuen Dimensionen gebildet wurden. Vielmehr vermischen sich die bestehenden Dimensionen miteinander und werden dadurch erweitert oder öffnen sich. Diese Erkenntnisse dienen auch zur Beantwortung der zweiten Forschungsfrage, welche neuen Dimensionen sich aus dem Word-Embedding ableiten lassen und welche Bedeutung diese haben. Die Hypothese, dass sich dabei keine neuen Dimensionen ableiten lassen, konnte bestätigt werden. Die aus dem Word-Embedding abgeleiteten Dimensionen sind nicht von Grund auf neu, enthalten aber Hinweise darauf, dass die RIASEC-Dimensionen eine geringere Trennschärfe besitzen als in der Theorie angenommen und weisen zudem auf komplexere Interessensräume hin. Die Erkenntnisse aus der Beantwortung der Forschungsfragen decken sich auch mit der Kritik am Modell, dass die Dimensionen des RIASEC-Modells nicht ausreichen um die Komplexität und Diversität der heutigen Berufswelt abzubilden (vgl. Armstrong et al., 2008).

5.3 Limitationen

Eine zentrale Limitation dieser Arbeit stellt das Word-Embedding dar, welches für das Vorhaben genutzt wurde. Die Repräsentation der ausgewählten Verben im Word-Embedding hängen stark von den Trainingsdaten ab. Abhängig davon, auf welcher Art Textdaten diese basierten, entstehen Unterschiede oder sogar Verzerrungen in der Bedeutungsstruktur. Es

besteht daher die Möglichkeit, dass Unterschiede, die für eine Zuordnung zu RIASEC-Dimensionen wichtig wären, durch das Modell unzureichend abgebildet wurden. Die Notwendigkeit zahlreiche Duplikate aus dem Word-Embedding zu entfernen, wirft zudem Fragen hinsichtlich dessen Qualität auf. Das Training eines eigenen Word-Embeddings mit Trainingsdaten, welche sich insbesondere auf Arbeitstätigkeiten und den zugehörigen Interessen beziehen, hätte diesen Problemen vorbeugen können. Jedoch ist das Training eines eigenen Word-Embeddings und die dazu notwendige Zusammensetzung von geeigneten Trainingsdaten sehr aufwändig.

Eine weitere Limitation zeigt sich in Bezug auf die Nutzung eines Word-Embeddings für das Vorhaben. Die Dimensionen eines Word-Embeddings sind nicht direkt interpretierbar und es lassen sich nicht gezielt Dimensionen der Semantik extrahieren. Durch die Dimensionsreduktion auf der Basis von Verben wurde diesem Problem zwar entgegengewirkt, jedoch kann nicht abschliessend bestimmt werden, ob sich diese Dimensionen auf rein semantische Unterschiede beziehen. Das schränkt die Interpretation der Faktoren ein, da die Zusammensetzung dieser durch Einflüsse von nicht semantischen Dimensionen verzerrt worden sein könnte. Hinzu kommt, dass Word-Embeddings Wörter anhand ihrer Verwendung in Kontexten repräsentieren und die Bedeutung von Verben stark kontextabhängig ist. Ein Beispiel dafür ist das Verb *malen* welches sowohl zur Dimension Artistic wie auch Realistic gehören könnte, je nachdem in welchem Kontext es genutzt wird.

Hinsichtlich der Hauptkomponentenanalyse zeigen sich weitere Limitationen. Durch die Reduktion der Dimensionen und Extraktion von einer geringen Anzahl an Faktoren reduzierte sich die erklärte Gesamtvarianz stark. Die Analyse der Ergebnisse und die Interpretation beziehen sich dabei vor allem auf die zehn stärksten positiven und negativen Faktorladungen. Dies erhöht zwar die Übersicht, beschränkt die Betrachtung jedoch auf die Extrembereiche. Eine weitere Einschränkung in Bezug auf die Interpretation ergibt sich aus der Berechnung der Mittelwerte. Die Standardabweichungen der Mittelwerte lagen in jedem

Fall über dem Mittelwert selbst, was auf eine hohe Streuung der Faktorladungen hinweist. Die Interpretation der Faktoren basierend auf den Mittelwerten ist daher vorsichtig zu betrachten.

Trotz der Limitationen bieten die Erkenntnisse erste wertvolle Hinweise darauf, wie KI mittels Word-Embedding, die menschliche Realität von Interessensstrukturmodellen versteht.

5.4 Ausblick

Für zukünftige Untersuchung wird die Entwicklung eines Word-Embeddings empfohlen, welches mit berufsspezifischen Textdaten trainiert wird. Dadurch könnte erreicht werden, dass sich die Wortrepräsentationen stärker durch berufsspezifische Dimensionen auszeichnen und so eine präzisere Abbildung von unterschiedlichen berufsbezogenen Faktoren ermöglichen.

Eine Weiterentwicklung des Vorgehens durch die Verwendung von komplexeren Sprachmodelle wie GPT könnte die Möglichkeit bieten, die kontextabhängige Bedeutung von Verben besser zu erfassen und differenziertere Faktoren zu extrahieren. Zusätzlich wäre es interessant, die Interpretation der Daten durch linguistisches Wissen zu ergänzen, um eine bessere Perspektive der semantischen Besonderheiten einzelner Verben zu erhalten.

In dieser Arbeit konnte zusammenfassend gezeigt werden, dass sich das RIASEC-Modell nicht eindeutig in den Daten eines Word-Embeddings widerspiegelt. Die Ergebnisse bieten jedoch interessante Hinweise auf Tendenzen, die dem Modell ähnlich sind. Eine Interpretation mit mehr Verben und einer Faktorenlösung mit mehr oder weniger Faktoren könnte daher interessante Ergebnisse und komplett neue Bedeutungsdimensionen mit sich bringen. Dadurch könnten auch neue Verben oder Dimensionen entdeckt werden, welche das Erfassen von beruflichen Interessen ergänzen und verbessern.

Literaturverzeichnis

- Armstrong, P. I., Day, S. X., McVay, J. P. & Rounds, J. (2008). Holland's RIASEC model as an integrative framework for individual differences. *Journal of Counseling Psychology*, 55(1), 1–18. <https://doi.org/10.1037/0022-0167.55.1.1>
- Armstrong, P. I., Smith, T. J., Donnay, D. A. C. & Rounds, J. (2004). The Strong Ring: A Basic Interest Model of Occupational Structure. *Journal of Counseling Psychology*, 51(3), 299–313. <https://doi.org/10.1037/0022-0167.51.3.299>
- Arseniev-Koehler, A. (2024). Theoretical Foundations and Limits of Word Embeddings: What Types of Meaning can They Capture? *Sociological Methods & Research*, 53(4), 1753–1793. <https://doi.org/10.1177/00491241221140142>
- Ascari, R., Giabelli, A., Malandri, L., Mercorio, F. & Mezzananza, M. (2024). A Fistful of Vectors: A Tool for Intrinsic Evaluation of Word Embeddings. *Cognitive Computation*, 16(3), 949–963. <https://doi.org/10.1007/s12559-023-10235-3>
- Bakker, J. D. (2024). *Applied Multivariate Statistics in R*. Seattle, WA: University of Washington.
- Bellman, R. (1978). *Artificial intelligence: can computers think?* Boston, MA: Boyd & Fraser Pub. Co.
- Bergmann, C. & Eder, F. (2005). AIST-R. Allgemeiner Interessen-Struktur-Test. In W. Sarges & H. Wottawa (Hrsg.), *Handbuch wirtschaftspsychologischer Testverfahren* (S. 53–58). Berlin: Pabst.
- BerufsbildungPlus.ch (n.d.). *Berufswahl*. Verfügbar unter: <https://www.berufsbildungplus.ch/Erwachsene/grundbildung/berufswahl>
- Bitkom & DFKI (2017). *Künstliche Intelligenz*. Positionspapier. Berlin: Bitkom und deutsches Forschungszentrum für künstliche Intelligenz (DFKI). Verfügbar unter:

https://www.dfki.de/fileadmin/user_upload/import/9744_171012-KI-Gipfelpapier-online.pdf

Bojanowski, P., Grave, E., Joulin, A. & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5, 135–146. https://doi.org/10.1162/tacl_a_00051

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V. & Kalai, A. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. *Advances in Neural Information Processing Systems*, 29. <https://doi.org/10.48550/arXiv.1607.06520>

Caliskan, A., Bryson, J. J. & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183–186. <https://doi.org/10.1126/science.aal4230>

Camacho-Collados, J. & Pilehvar, M. T. (2018). From Word To Sense Embeddings: A Survey on Vector Representations of Meaning. *Journal of Artificial Intelligence Research*, 63, 743–788. <https://doi.org/10.1613/jair.1.11259>

Charniak, E. & McDermott, D. V. (1985). *Introduction to artificial intelligence*. Reading, MA: Addison-Wesley.

Deepset (2018). Deepset German Word2Vec Embedding. *German Word-Embeddings*. Verfügbar unter: <https://www.deepset.ai/german-word-embeddings>

Deepset (2023). Deepset German GloVe Embedding. *German Word-Embeddings*. Verfügbar unter: <https://www.deepset.ai/german-word-embeddings>

Eder, F. & Bergmann, C. (2015). Das Person-Umwelt-Modell von J. L. Holland: Grundlagen - Konzepte - Anwendungen. In C. Tarnai & F.G. Hartmann (Hrsg.), *Berufliche Interessen. Beiträge zur Theorie von J.L. Holland* (S. 11–30). Münster: Waxmann Verlag.

- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists* (Multivariate applications book series). Mahwah, NJ: L. Erlbaum Associates.
- Faruqui, M., Tsvetkov, Y., Rastogi, P. & Dyer, C. (2016). Problems With Evaluation of Word Embeddings Using Word Similarity Tasks. In *Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP* (pp. 30–35). Berlin: Association for Computational Linguistics. <https://doi.org/10.18653/v1/W16-2506>
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). London: SAGE.
- Gfrörer, T., Stoll, G., Rieger, S., Trautwein, U. & Nagengast, B. (2022). The Development of Vocational Interests in Early Adolescence: Stability, Change, and State-Trait Components. *European Journal of Personality*, 36(6), 942–964. <https://doi.org/10.1177/08902070211035630>
- Glatzeder, B., Goel, V. & Müller, A. (Eds.). (2010). *Towards a Theory of Thinking*. Berlin: Springer. <https://doi.org/10.1007/978-3-642-03129-8>
- Göbel, A. (2018). Liste Verben alphabetisch. *Verblisten*. Verfügbar unter: <https://www.verblisten.de/listen/verben/anfangsbuchstabe/ueberblick>
- Goodfellow, I., Courville, A. & Bengio, Y. (2016). *Deep learning*. Cambridge, MA: MIT Press.
- Grave, E., Bojanowski, P., Gupta, P., Joulin, A. & Mikolov, T. (2018). Learning Word Vectors for 157 Languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki: European Language Resources Association (ELRA). <https://doi.org/10.48550/arXiv.1802.06893>
- Haenlein, M. & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, 61(4), 5–14. <https://doi.org/10.1177/0008125619864925>

- Hansen, J.-I. C. (2013). Nature, importance and assessment of interests. In S.D. Brown & R.W. Lent (Eds.), *Career development and counseling: putting theory and research to work* (2nd ed., pp. 387–416). Hoboken, NJ: Wiley.
- Hartmann, F., Heine, J.-H. & Ertl, B. (2021). Concepts and Coefficients Based on John L. Holland's Theory of Vocational Choice Examining the R Package holland. *Psych*, 3(4), 728–750. <https://doi.org/10.3390/psych3040047>
- Haynes, J.-D. (2012). Brain reading. In S. Richmond, G. Rees & S. Edwards (Eds.), *I Know What You're Thinking: Brain Imaging and Mental Privacy* (pp. 29–40). Oxford: Oxford University Press.
- Hell, B., Pässler, K. & Schuler, H. (2009). „was-studiere-ich.de“ - Konzept, Nutzen und Anwendungsmöglichkeiten. *Zeitschrift für Studium und Beratung*, 4(1), 9–14.
- Hidi, S. (1990). Interest and Its Contribution as a Mental Resource for Learning. *Review of Educational Research*, 60(4), 549–571. <https://doi.org/10.3102/00346543060004549>
- Hidi, S. & Renninger, K. A. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist*, 41(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4
- Hirschi, A. (2018). The Fourth Industrial Revolution: Issues and Implications for Career Research and Practice. *The Career Development Quarterly*, 66(3), 192–204. <https://doi.org/10.1002/cdq.12142>
- Hirschle, J. (2022). *Deep Natural Language Processing: Einstieg in Word Embedding, Sequence-to-Sequence-Modelle und Transformers mit Python*. München: Hanser. <https://doi.org/10.3139/9783446473904>
- Hogan, R. & Blake, R. (1999). John Holland's Vocational Typology and Personality Theory. *Journal of Vocational Behavior*, 55(1), 41–56. <https://doi.org/10.1006/jvbe.1999.1696>
- Holland, J. L. (1997). *Making vocational choices: a theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

- Iglewicz, B. & Hoaglin, D. C. (1993). *How to detect and handle outliers*. Milwaukee, WI: ASQC Quality Press.
- Jolliffe, I. T. & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065). The Royal Society.
<https://doi.org/10.1098/rsta.2015.0202>
- Jörin, S., Stoll, F., Bergmann, C. & Eder, F. (2003). Explorix. Deutschsprachiger Self-Directed Search. In W. Sarges & H. Wottawa (Hrsg.), *Handbuch wirtschaftspsychologischer Testverfahren* (S. 309–316). Berlin: Pabst.
- Joulin, A., Grave, E., Bojanowski, P. & Mikolov, T. (2017). Bag of Tricks for Efficient Text Classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 427–431). Valencia: Association for Computational Linguistics. <https://doi.org/10.48550/arXiv.1607.01759>
- Juntunen, C. L., Motl, T. C. & Rozzi, M. (2019). Major Career Theories: International and Developmental Perspectives. In J.A. Athanasou & H.N. Perera (Eds.), *International Handbook of Career Guidance* (pp. 45–72). Cham: Springer International Publishing.
https://doi.org/10.1007/978-3-030-25153-6_3
- Jurafsky, D. & Martin, J. H. (2014). *Speech and language processing* (2nd ed.). Harlow: Pearson.
- Kaplan, J. (2017). *Künstliche Intelligenz*. Frechen: mitp Verlags GmbH & Co. KG.
- Köper, M., Scheible, C. & Schulte im Walde, S. (2015). Multilingual Reliability and “Semantic” Structure of Continuous Word Spaces. *11th International Conference on Computational Semantics* (pp. 40–45). London: Association for Computational Linguistics.

- Kozlowski, A. C., Taddy, M. & Evans, J. A. (2019). The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review*, 84(5), 905–949. <https://doi.org/10.1177/0003122419877135>
- Krapp, A. (2002). Structural and dynamic aspects of interest development: theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, 12(4), 383–409. [https://doi.org/10.1016/S0959-4752\(01\)00011-1](https://doi.org/10.1016/S0959-4752(01)00011-1)
- Krapp, A. & Prenzel, M. (2011). Research on Interest in Science: Theories, methods, and findings. *International Journal of Science Education*, 33(1), 27–50. <https://doi.org/10.1080/09500693.2010.518645>
- Kuckartz, U., Rädiker, S., Ebert, T. & Schehl, J. (2013). *Statistik: Eine verständliche Einführung* (2., überarbeitete Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-19890-3>
- Kuhn, D. (2000). Metacognitive Development. *Current Directions in Psychological Science*, 9(5), 178–181. <https://doi.org/10.1111/1467-8721.00088>
- Kurzweil, R. (1990). *The age of intelligent machines*. Cambridge, MA: MIT Press.
- Lane, H. & Dyschel, M. (2025). *Natural language processing in action* (2nd ed.). Shelter Island, NY: Manning Publications.
- Lee, K.-F. (2018). *AI superpowers: China, Silicon Valley, and the new world order*. Boston, MA: Houghton Mifflin Harcourt.
- Leviant, I. & Reichart, R. (2015). Separated by an Un-common Language: Towards Judgment Language Informed Vector Space Modeling. arXiv. <https://doi.org/10.48550/arXiv.1508.00106>
- Levy, O. & Goldberg, Y. (2014). Linguistic Regularities in Sparse and Explicit Word Representations. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning* (pp. 171–180). Ann Arbor, MI: Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-1618>

- Mandera, P., Keuleers, E. & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, 92, 57–78. <https://doi.org/10.1016/j.jml.2016.04.001>
- McCarthy, J., Minsky, M. L., Rochester, N. & Shannon, C. E. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Research Proposal. Hanover, NH: Dartmouth College. Verfügbar unter: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>
- Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv. <https://doi.org/10.48550/arXiv.1301.3781>
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems*, 26. <https://doi.org/10.48550/arXiv.1310.4546>
- Mikolov, T., Yih, W. & Zweig, G. (2013). Linguistic Regularities in Continuous Space Word Representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 746–751). Atlanta, GA: Association for Computational Linguistics.
- Mitchell, M. (2019). *Artificial intelligence: a guide for thinking humans*. New York, NY: Farrar, Straus and Giroux.
- Mount, M. K., Barrick, M. R., Scullen, S. M. & Rounds, J. (2005). Higher-order dimensions of the Big Five personality traits and the Big Six vocational interests Types. *Personnel Psychology*, 58(2), 447–478. <https://doi.org/10.1111/j.1744-6570.2005.00468.x>
- Müller, A. (2015). *Analyse von Wort-Vektoren deutscher Textkorpora*. Bachelorarbeit. Berlin: Technische Universität Berlin. Verfügbar unter: <https://devmount.github.io/GermanWordEmbeddings/>

- Naseem, U., Razzak, I., Khan, S. K. & Prasad, M. (2021). A Comprehensive Survey on Word Representation Models: From Classical to State-of-the-Art Word Representation Language Models. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 20(5), 1–35. <https://doi.org/10.1145/3434237>
- Nauta, M. M. (2010). The development, evolution, and status of Holland's theory of vocational personalities: Reflections and future directions for counseling psychology. *Journal of Counseling Psychology*, 57(1), 11–22. <https://doi.org/10.1037/a0018213>
- Neumayer, S., Nimmer, M., Setzer, S. & Steidl, G. (2020). On the Robust PCA and Weiszfeld's Algorithm. *Applied Mathematics & Optimization*, 82(3), 1017–1048. <https://doi.org/10.1007/s00245-019-09566-1>
- O*NET OnLine (2025). *O*NET OnLine Help. OnLine Overview*. Verfügbar unter: <https://www.onetonline.org/help/online/>
- Osterlind, S. J. & Everson, H. T. (2009). *Differential item functioning* (2nd ed.). Thousand Oaks, CA: SAGE.
- Paass, G. & Hecker, D. (2024). *Artificial Intelligence: What Is Behind the Technology of the Future?* Cham: Springer International Publishing AG.
- Pennington, J., Socher, R. & Manning, C. (2014). Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). Doha: Association for Computational Linguistics. <https://doi.org/10.3115/v1/D14-1162>
- Poole, D. L., Mackworth, A. K. & Goebel, R. (1998). *Computational intelligence: a logical approach*. Oxford, NY: Oxford University Press.
- Prediger, D. J. (1982). Dimensions underlying Holland's hexagon: Missing link between interests and occupations? *Journal of Vocational Behavior*, 21(3), 259–287. [https://doi.org/10.1016/0001-8791\(82\)90036-7](https://doi.org/10.1016/0001-8791(82)90036-7)

- Prediger, D. J. & Cole, N. S. (1975). Sex-role socialization and employment realities: Implications for vocational interest measures. *Journal of Vocational Behavior*, 7(2), 239–251. [https://doi.org/10.1016/0001-8791\(75\)90064-0](https://doi.org/10.1016/0001-8791(75)90064-0)
- Reeve, B. B. & Mâsse, L. C. (2004). Item Response Theory Modeling for Questionnaire Evaluation. In S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin et al. (Eds.), *Methods for Testing and Evaluating Survey Questionnaires* (pp. 247–273). Hoboken, NJ: Wiley. <https://doi.org/10.1002/0471654728.ch13>
- Renninger, K. A. & Hidi, S. E. (2019). Interest Development and Learning. In K.A. Renninger & S.E. Hidi (Eds.), *The Cambridge Handbook of Motivation and Learning* (pp. 265–290). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316823279>
- Roch, S. (2025). *Mathematical Methods in Data Science: Bridging Theory and Applications with Python*. Cambridge: Cambridge University Press.
- Rong, X. (2014). word2vec Parameter Learning Explained. arXiv. <https://doi.org/10.48550/ARXIV.1411.2738>
- Rounds, J. & Tracey, T. J. (1993). Prediger's dimensional representation of Holland's RIASEC circumplex. *Journal of Applied Psychology*, 78(6), 875–890. <https://doi.org/10.1037/0021-9010.78.6.875>
- Rounds, J., Tracey, T. J. & Hubert, L. (1992). Methods for evaluating vocational interest structural hypotheses. *Journal of Vocational Behavior*, 40(2), 239–259. [https://doi.org/10.1016/0001-8791\(92\)90073-9](https://doi.org/10.1016/0001-8791(92)90073-9)
- Russell, S. J. & Norvig, P. (2021). *Artificial intelligence: a modern approach* (4th ed.). Hoboken, NJ: Pearson.
- Schakel, A. M. J. & Wilson, B. J. (2015). Measuring Word Significance using Distributed Representations of Words. arXiv. <https://doi.org/10.48550/arXiv.1508.02297>

- Schmidt-Atzert, L., Krumm, S. & Amelang, M. (Hrsg.). (2021). *Psychologische Diagnostik*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-61643-7>
- scikit-learn (n.d.). *PCA*. Verfügbar unter: <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA>
- Sharf, R. S. (2013). Advances in Theories of Career Development. In W.B. Walsh, M.L. Savickas & P.J. Hartung (Eds.), *Handbook of vocational psychology: theory, research, and practice* (4th ed., pp. 3–32). New York, NY: Routledge.
- Silvia, P. J. (2006). *Exploring the psychology of interest*. Oxford, NY: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195158557.001.0001>
- Sodano, S. M. (2015). Meaning, measurement, and assessment of vocational interests for career intervention. In P.J. Hartung, M.L. Savickas & W.B. Walsh (Eds.), *APA handbook of career intervention, Volume 1: Foundations*. (pp. 281–301). Washington, WA: American Psychological Association. <https://doi.org/10.1037/14438-016>
- Sowmya, V. B., Majumder, B., Gupta, A. & Surana, H. (2020). *Practical natural language processing: a comprehensive guide to building real-world NLP systems*. Sebastopol, CA: O'Reilly Media.
- Spokane, A. R., Luchetta, E. J. & Richwine, M. H. (2002). Holland's Theory of Personalities in Work Environments. *Career Choice and Development*, 15(1), 373–426.
- Su, R., Rounds, J. & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, 135(6), 859–884. <https://doi.org/10.1037/a0017364>
- Su, R., Tay, L., Liao, H.-Y., Zhang, Q. & Rounds, J. (2019). Toward a dimensional model of vocational interests. *Journal of Applied Psychology*, 104(5), 690–714. <https://doi.org/10.1037/apl0000373>

- Toshevskva, M., Stojanovska, F. & Kalajdjieski, J. (2020). Comparative Analysis of Word Embeddings for Capturing Word Similarities. In *Proceedings of the 6th International Conference on Natural Language Processing* (pp. 9–24). Copenhagen: NATP.
<https://doi.org/10.5121/csit.2020.100402>
- Treistman, A., Mughaz, D., Stulman, A. & Dvir, A. (2022). Word embedding dimensionality reduction using dynamic variance thresholding (DyVaT). *Expert Systems with Applications*, 208(118157). <https://doi.org/10.1016/j.eswa.2022.118157>
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, LIX(236), 433–460.
<https://doi.org/10.1093/mind/LIX.236.433>
- Vargas, F. & Cotterell, R. (2020). Exploring the Linear Subspace Hypothesis in Gender Bias Mitigation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 2902–2913). Association for Computational Linguistics. <https://doi.org/10.48550/arXiv.2009.09435>
- Wang, B., Wang, A., Chen, F., Wang, Y. & Kuo, C.-C. J. (2019). Evaluating word embedding models: methods and experimental results. *APSIPA Transactions on Signal and Information Processing*, 8(1). <https://doi.org/10.1017/ATSIP.2019.12>
- Wang, P., Yan, Y., Tian, Y., Feng, X., Zhou, H., Wang, H. et al. (2024). Contemporary Chinese vocational interest scale in the digital age: development and validation. *Humanities and Social Sciences Communications*, 11(1), 1405.
<https://doi.org/10.1057/s41599-024-03923-4>
- Wetzel, E., Hell, B. & Pässler, K. (2012). Comparison of Different Test Construction Strategies in the Development of a Gender Fair Interest Inventory Using Verbs. *Journal of Career Assessment*, 20(1), 88–104.
<https://doi.org/https://doi.org/10.1177/106907271141716>
- Wooldridge, M. J. (Eds.). (2000). *Reasoning about rational agents*. Cambridge, MA: MIT Press.

Xue, J., Wang, Y.-C., Wei, C. & Kuo, C.-C. J. (2024). Word Embedding Dimension Reduction via Weakly-Supervised Feature Selection. arXiv.

<https://doi.org/10.48550/arXiv.2407.12342>

Zhao, J., Wang, T., Yatskar, M., Cotterell, R., Ordonez, V. & Chang, K.-W. (2019). Gender Bias in Contextualized Word Embeddings. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 629–634). Minneapolis, MN: Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1064>

Abbildungsverzeichnis

| | |
|--|----|
| Abbildung 1. Illustration zur Funktionsweise von Word-Embeddings in einem zweidimensionalen Raum (eigene Darstellung)..... | 17 |
| Abbildung 2. Hexagon des RIASEC-Modells nach Holland (1997) (eigene Darstellung) | 26 |
| Abbildung 3. RIASEC-Modell mit Predigers (1982) zugrundeliegenden Dimensionen (eigene Darstellung)..... | 30 |
| Abbildung 4. RIASEC-Modell mit Hogans (1983, zitiert nach Sodano, 2015, S. 284) zugrundeliegenden Dimensionen (eigene Darstellung) | 31 |
| Abbildung 5. Screeplot zur Bestätigung der Faktorenanzahl bei der Hauptkomponentenanalyse. | 46 |

Tabellenverzeichnis

| | |
|--|----|
| Tabelle 1 Verben aller VIT-Versionen mit zugehörigen RIASEC-Dimensionen (Wetzel et al., 2012)..... | 37 |
| Tabelle 2 Als Ausreisser identifizierte Verben mit zugeordneter RIASEC-Dimension | 45 |
| Tabelle 3 Anzahl enthaltener Verben aus dem VIT in den Word-Embeddings..... | 48 |
| Tabelle 4 Ergebnisse der intrinsischen Evaluation..... | 48 |
| Tabelle 5 Mittelwerte der RIASEC-Dimensionen innerhalb der Faktoren..... | 51 |
| Tabelle 6 Abkürzungen der Achsenausprägungen auf den zugrundeliegenden Achsen des RIASEC-Modells | 53 |
| Tabelle 7 Top-10-Verben der 6-Faktorenlösung für Faktor 1..... | 54 |
| Tabelle 8 Bottom-10-Verben der 6-Faktorenlösung für Faktor 1 | 55 |
| Tabelle 9 Top-10-Verben der 6-Faktorenlösung für Faktor 2..... | 56 |
| Tabelle 10 Bottom-10-Verben der 6-Faktorenlösung für Faktor 2 | 57 |
| Tabelle 11 Top-10-Verben der 6-Faktorenlösung für Faktor 3..... | 59 |
| Tabelle 12 Bottom-10-Verben der 6-Faktorenlösung für Faktor 3 | 60 |
| Tabelle 13 Top-10-Verben der 6-Faktorenlösung für Faktor 4..... | 61 |
| Tabelle 14 Bottom-10-Verben der 6-Faktorenlösung für Faktor 4 | 62 |
| Tabelle 15 Top-10-Verben der 6-Faktorenlösung für Faktor 5..... | 63 |
| Tabelle 16 Bottom-10-Verben der 6-Faktorenlösung für Faktor 5 | 64 |
| Tabelle 17 Top-10-Verben der 6-Faktorenlösung für Faktor 6..... | 65 |
| Tabelle 18 Bottom-10-Verben der 6-Faktorenlösung für Faktor 6 | 67 |

Hilfsmittelverzeichnis mit Verwendungszweck

| KI-Assistenzsystem | Teile / Stelle(n) in der Arbeit | Einsatz |
|---------------------------|--|---|
| ChatGPT | Gesamte Arbeit | Hilfe zu Formulierungen, Überprüfung von Grammatik & Rechtschreibung, Überprüfung auf Vollständigkeit, prüfen ob einzelne Inhalte sinngemäss korrekt verstanden wurden, Übersetzung von unverständlichen Quellen, Literaturrecherche. |
| DeepL | Abstract | Hilfe zur Übersetzung ins Englische |
| ChatGPT | Abstract | Kürzung Abstract auf vorgegebene Zeichenzahl, Hilfe zur Übersetzung ins Englische |
| ChatGPT | Theorie | Erklärung von unbekanntem technischen Konzepten. |
| ChatGPT | Methodik | Hilfe zur Erstellung der Codelines für Jupyter Notebook und Troubleshooting. |

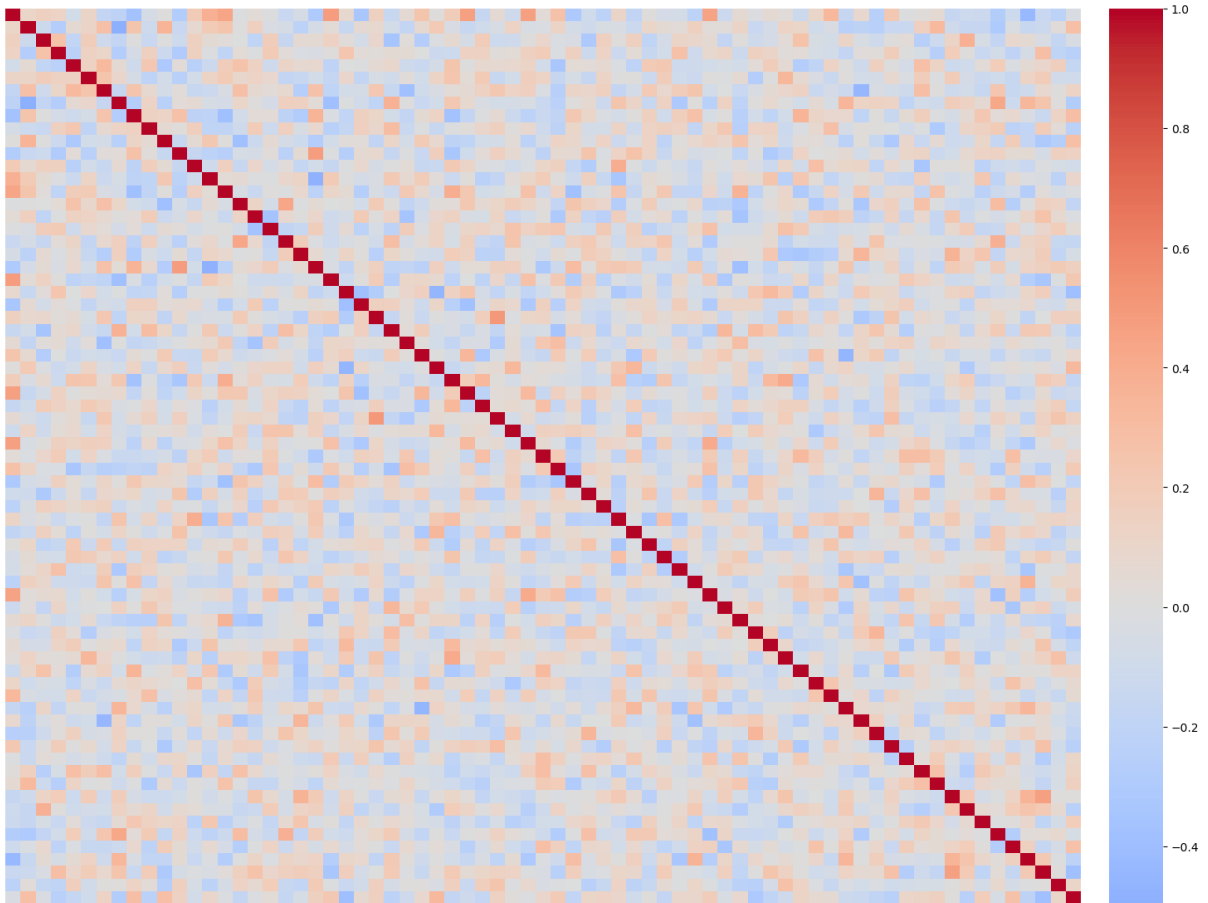
Anhang

Anhang A: Zusammensetzung der drei Versionen des VITs

| Holland-Dimension | Verb | Classic | Balanced | IRT |
|-------------------|-----------------|---------|----------|-----|
| Realistic | aufbauen | X | | |
| Realistic | schnitzen | X | X | X |
| Realistic | zerlegen | | X | X |
| Realistic | umgraben | | X | |
| Realistic | bohren | X | X | X |
| Realistic | kleben | X | X | X |
| Realistic | hämmern | X | | X |
| Realistic | pflanzen | X | X | X |
| Realistic | reparieren | X | X | X |
| Realistic | schrauben | X | | |
| Realistic | tapezieren | | X | X |
| Investigative | vermuten | | X | X |
| Investigative | folgern | X | | X |
| Investigative | entdecken | X | X | |
| Investigative | finden | X | X | X |
| Investigative | erfinden | X | | |
| Investigative | nachforschen | X | | X |
| Investigative | beobachten | | X | X |
| Investigative | rätseln | | X | |
| Investigative | erkennen | X | X | |
| Investigative | auskundschaften | | X | X |
| Investigative | testen | | | X |
| Investigative | ausdenken | X | | |
| Investigative | probieren | X | X | X |
| Artistic | gestalten | X | X | X |
| Artistic | malen | X | X | |
| Artistic | ausstellen | | X | X |
| Artistic | verfilmen | X | X | X |
| Artistic | reimen | | | X |
| Artistic | singen | X | | |
| Artistic | übersetzen | | | X |
| Artistic | dichten | X | X | |
| Social | begleiten | X | | X |
| Social | betreuen | X | | X |
| Social | beistehen | X | | X |
| Social | trösten | X | | |
| Social | beraten | | X | |
| Social | ausbilden | | X | X |
| Social | einfühlen | X | | |
| Social | beschützen | | X | X |
| Social | helfen | X | X | |
| Social | belehren | | X | X |
| Social | lehren | | X | |

| | | | | |
|--------------|-----------------------|---|---|---|
| Social | zuhören | | | X |
| Social | behüten | X | X | X |
| Social | erziehen | X | | |
| Social | jdm. etwas beibringen | | X | |
| Enterprising | erkämpfen | | X | |
| Enterprising | handeln | | X | X |
| Enterprising | durchgreifen | X | | X |
| Enterprising | überzeugen | X | X | |
| Enterprising | führen | X | | |
| Enterprising | durchsetzen | X | | X |
| Enterprising | beeinflussen | X | | X |
| Enterprising | investieren | | | X |
| Enterprising | leiten | X | X | |
| Enterprising | verhandeln | X | | |
| Enterprising | überreden | | X | |
| Enterprising | anwerben | | X | X |
| Enterprising | auffordern | X | X | X |
| Enterprising | fordern | | X | X |
| Conventional | verwalten | X | X | X |
| Conventional | beantragen | X | X | X |
| Conventional | kalkulieren | X | | X |
| Conventional | kontrollieren | | X | X |
| Conventional | zählen | X | | |
| Conventional | einordnen | X | X | X |
| Conventional | nummerieren | X | X | |
| Conventional | aufzählen | X | X | X |
| Conventional | ordnen | X | | |
| Conventional | aussortieren | | X | X |
| Conventional | überweisen | | X | X |

Anhang B: Korrelationsmatrix des reduzierten Word-Embeddings als Heatmap



Die Korrelation der Dimensionen wird farblich dargestellt, und jedes Feld entspricht einer Dimension. Die negativen Korrelationen werden in Blautönen, die positiven Korrelationen in Rottönen dargestellt. Je näher der Korrelationswert bei 0 ist desto blasser werden die Farben.

Anhang C: Übersichtstabelle der in den Word-Embeddings fehlenden Verben

| BA-Word2Vec | Deepset-Word2Vec | Deepset-GloVe | WE-fastText |
|-----------------------|-----------------------|-----------------------|-----------------------|
| hämmern | hämmern | jdm. etwas beibringen | jdm. etwas beibringen |
| rätseln | rätseln | | |
| übersetzen | übersetzen | | |
| trösten | trösten | | |
| einfühlen | einfühlen | | |
| beschützen | beschützen | | |
| zuhören | zuhören | | |
| behüten | behüten | | |
| jdm. etwas beibringen | jdm. etwas beibringen | | |
| erkämpfen | erkämpfen | | |
| überzeugen | überzeugen | | |
| führen | führen | | |
| überreden | überreden | | |
| zählen | zählen | | |
| aufzählen | aufzählen | | |
| überweisen | überweisen | | |

Anhang D: Erklärte Varianz der Hauptkomponentenanalyse pro Faktor

| Komponente | Anfängliche Eigenwerte | | | Rotierte Summe der quadrierten Ladungen |
|------------|------------------------|---------------|--------------|---|
| | Gesamt | % der Varianz | Kumulierte % | Gesamt |
| 1 | 0.01953861 | 12.18239221 | 12.18239221 | 1.0061393 |
| 2 | 0.01373033 | 8.56090817 | 20.74330038 | 1.0086631 |
| 3 | 0.01172971 | 7.31351458 | 28.05681496 | 1.0132384 |
| 4 | 0.01006946 | 6.27834601 | 34.33516097 | 1.0268971 |
| 5 | 0.00825339 | 5.14601792 | 39.48117889 | 1.0105678 |
| 6 | 0.00792162 | 4.93915614 | 44.42033503 | 1.0363298 |
| 7 | 0.00661387 | 4.12377363 | 48.54410866 | |
| 8 | 0.00586783 | 3.65861023 | 52.20271889 | |
| 9 | 0.00526373 | 3.28195505 | 55.48467394 | |
| 10 | 0.00523241 | 3.2624234 | 58.74709734 | |
| 11 | 0.00504173 | 3.14353475 | 61.89063209 | |
| 12 | 0.00449273 | 2.80123196 | 64.69186405 | |
| 13 | 0.00406479 | 2.53441108 | 67.22627513 | |
| 14 | 0.0037506 | 2.33850934 | 69.56478447 | |
| 15 | 0.00342081 | 2.13288932 | 71.6976738 | |
| 16 | 0.00311775 | 1.94392707 | 73.64160086 | |
| 17 | 0.00297838 | 1.85702857 | 75.49862944 | |
| 18 | 0.002946 | 1.83684073 | 77.33547017 | |
| 19 | 0.00272259 | 1.69754735 | 79.03301752 | |
| 20 | 0.00248124 | 1.54705886 | 80.58007638 | |
| 21 | 0.0023187 | 1.44571733 | 82.02579371 | |
| 22 | 0.00213071 | 1.3285022 | 83.35429591 | |
| 23 | 0.00206787 | 1.28932276 | 84.64361866 | |
| 24 | 0.00190077 | 1.18513468 | 85.82875335 | |
| 25 | 0.00172917 | 1.07814156 | 86.90689491 | |
| 26 | 0.00157887 | 0.98442836 | 87.89132326 | |
| 27 | 0.00150912 | 0.94094389 | 88.83226715 | |
| 28 | 0.00143209 | 0.89291568 | 89.72518283 | |
| 29 | 0.00134525 | 0.83876949 | 90.56395233 | |
| 30 | 0.00123872 | 0.77234507 | 91.3362974 | |
| 31 | 0.00121217 | 0.75579171 | 92.09208911 | |
| 32 | 0.0011566 | 0.72114688 | 92.81323599 | |
| 33 | 0.00103514 | 0.64541345 | 93.45864944 | |
| 34 | 0.00096839 | 0.60379654 | 94.06244598 | |
| 35 | 0.00093495 | 0.58294647 | 94.64539245 | |
| 36 | 0.00087949 | 0.54836462 | 95.19375707 | |

| | | | |
|----|------------|------------|-------------|
| 37 | 0.0007154 | 0.44605423 | 95.6398113 |
| 38 | 0.00067547 | 0.4211587 | 96.06097 |
| 39 | 0.00062407 | 0.38910933 | 96.45007933 |
| 40 | 0.00059448 | 0.37066108 | 96.82074041 |
| 41 | 0.00054269 | 0.33836996 | 97.15911037 |
| 42 | 0.00050166 | 0.31278709 | 97.47189746 |
| 43 | 0.00044409 | 0.27688872 | 97.74878618 |
| 44 | 0.00039866 | 0.24856877 | 97.99735494 |
| 45 | 0.00037168 | 0.23174105 | 98.22909599 |
| 46 | 0.00032854 | 0.20484346 | 98.43393945 |
| 47 | 0.00029176 | 0.18191168 | 98.61585113 |
| 48 | 0.00027451 | 0.17115619 | 98.78700732 |
| 49 | 0.00026757 | 0.16683092 | 98.95383824 |
| 50 | 0.00023794 | 0.14835752 | 99.10219577 |
| 51 | 0.00022704 | 0.14155883 | 99.2437546 |
| 52 | 0.00018385 | 0.11463009 | 99.35838469 |
| 53 | 0.00015748 | 0.09819066 | 99.45657535 |
| 54 | 0.00015604 | 0.09729204 | 99.55386739 |
| 55 | 0.00013261 | 0.08268087 | 99.63654826 |
| 56 | 0.00010595 | 0.06606021 | 99.70260847 |
| 57 | 0.00009215 | 0.05745476 | 99.76006324 |
| 58 | 0.00008123 | 0.05064431 | 99.81070754 |
| 59 | 0.00007349 | 0.04581836 | 99.85652591 |
| 60 | 0.00006299 | 0.03927701 | 99.89580292 |
| 61 | 0.00004224 | 0.02633807 | 99.92214099 |
| 62 | 0.00003593 | 0.02240471 | 99.9445457 |
| 63 | 0.00002874 | 0.01791769 | 99.96246339 |
| 64 | 0.0000201 | 0.01253263 | 99.97499602 |
| 65 | 0.00001451 | 0.00904963 | 99.98404565 |
| 66 | 0.00001195 | 0.00744897 | 99.99149462 |
| 67 | 0.00000818 | 0.00509879 | 99.99659341 |
| 68 | 0.0000045 | 0.00280322 | 99.99939663 |
| 69 | 0.00000084 | 0.00052484 | 99.99992148 |
| 70 | 0.00000013 | 0.00007852 | 100 |
| 71 | 0 | 0 | 100 |

Extraktionsmethode: Hauptkomponentenanalyse.

Anmerkung. Der Eigenwert der 71. Komponente beträgt 3.845×10^{-33} . Der zugehörige Anteil an der Gesamtvarianz liegt bei $2.397 \times 10^{-32}\%$. Beide Werte werden in der Tabelle auf Grund von Rundungen mit 0 dargestellt.

Anhang E: Vollständige Ladungsmatrix der Hauptkomponentenanalyse

| | | Faktorladungen | | | | | |
|---------------|-----------------|----------------|----------|----------|----------|----------|----------|
| RIASEC | Verb | Faktor 1 | Faktor 2 | Faktor 3 | Faktor 4 | Faktor 5 | Faktor 6 |
| Realistic | aufbauen | -0.073 | -0.096 | 0.155 | 0.104 | -0.275 | -0.074 |
| Realistic | bohren | -0.001 | 0.027 | -0.036 | -0.017 | -0.067 | 0.131 |
| Realistic | hämmern | 0.016 | 0.107 | -0.019 | -0.011 | 0.023 | -0.123 |
| Realistic | kleben | -0.062 | -0.009 | 0.090 | 0.168 | -0.003 | 0.035 |
| Realistic | pflanzen | 0.245 | -0.067 | 0.170 | 0.044 | -0.103 | -0.011 |
| Realistic | reparieren | 0.086 | 0.050 | -0.106 | -0.210 | 0.051 | 0.091 |
| Realistic | schnitzen | 0.008 | -0.008 | -0.062 | 0.020 | -0.050 | -0.036 |
| Realistic | schrauben | 0.150 | 0.009 | -0.042 | 0.046 | -0.090 | -0.035 |
| Realistic | tapezieren | -0.037 | -0.109 | 0.110 | -0.072 | -0.070 | -0.009 |
| Realistic | umgraben | 0.008 | 0.099 | -0.073 | 0.004 | -0.132 | -0.097 |
| Realistic | zerlegen | -0.089 | 0.067 | 0.003 | 0.073 | -0.078 | -0.004 |
| Investigative | ausdenken | -0.006 | 0.002 | 0.048 | -0.057 | 0.162 | 0.138 |
| Investigative | auskundschaften | -0.023 | -0.011 | 0.081 | 0.037 | -0.089 | -0.108 |
| Investigative | beobachten | -0.037 | -0.103 | -0.005 | -0.034 | -0.273 | -0.162 |
| Investigative | entdecken | -0.026 | 0.136 | 0.211 | 0.291 | 0.226 | 0.069 |
| Investigative | erfinden | -0.001 | 0.031 | -0.038 | 0.085 | 0.024 | 0.043 |
| Investigative | erkennen | 0.051 | -0.112 | 0.009 | 0.112 | 0.030 | 0.559 |
| Investigative | finden | 0.769 | -0.001 | 0.002 | -0.006 | 0.007 | 0.011 |
| Investigative | folgern | 0.113 | -0.048 | 0.036 | -0.053 | -0.098 | -0.056 |
| Investigative | nachforschen | -0.058 | -0.031 | -0.004 | 0.083 | 0.066 | -0.010 |
| Investigative | probieren | -0.017 | -0.024 | -0.026 | -0.036 | 0.267 | 0.245 |

| | | | | | | | |
|---------------|--------------|--------|--------|--------|--------|--------|--------|
| Investigative | rätseln | -0.046 | -0.083 | 0.022 | 0.086 | -0.021 | -0.119 |
| Investigative | testen | -0.066 | 0.224 | -0.253 | -0.064 | 0.092 | 0.048 |
| Investigative | vermuten | -0.021 | -0.105 | -0.037 | -0.053 | 0.142 | -0.029 |
| Artistic | ausstellen | 0.026 | 0.035 | 0.023 | -0.150 | -0.062 | 0.063 |
| Artistic | dichten | -0.007 | 0.063 | -0.014 | 0.043 | -0.083 | -0.089 |
| Artistic | gestalten | -0.049 | 0.182 | 0.359 | 0.025 | -0.144 | 0.221 |
| Artistic | malen | 0.032 | -0.023 | 0.186 | 0.160 | 0.237 | -0.136 |
| Artistic | reimen | -0.014 | -0.048 | -0.071 | -0.123 | -0.057 | -0.010 |
| Artistic | singen | 0.112 | -0.002 | 0.141 | 0.061 | -0.271 | -0.001 |
| Artistic | übersetzen | -0.099 | 0.024 | -0.050 | -0.056 | -0.079 | 0.012 |
| Artistic | verfilmen | -0.055 | -0.046 | -0.009 | -0.079 | 0.060 | 0.054 |
| Social | ausbilden | -0.051 | 0.050 | 0.023 | -0.037 | 0.032 | 0.070 |
| Social | begleiten | 0.178 | 0.112 | 0.083 | 0.100 | 0.209 | -0.245 |
| Social | behüten | -0.027 | 0.016 | -0.075 | 0.029 | -0.013 | -0.066 |
| Social | beistehen | 0.078 | -0.104 | 0.101 | 0.031 | -0.021 | 0.017 |
| Social | belehren | -0.051 | 0.009 | -0.025 | -0.077 | 0.062 | 0.047 |
| Social | beraten | -0.137 | 0.126 | -0.034 | -0.008 | -0.308 | 0.055 |
| Social | beschützen | 0.037 | -0.065 | 0.028 | 0.093 | -0.023 | -0.247 |
| Social | betreuen | 0.009 | -0.041 | -0.117 | 0.014 | 0.064 | -0.038 |
| Social | einfühlen | -0.135 | -0.034 | -0.009 | 0.034 | 0.085 | -0.006 |
| Social | erziehen | -0.023 | 0.015 | 0.101 | 0.021 | 0.100 | -0.089 |
| Social | helfen | -0.009 | 0.746 | 0.023 | -0.010 | 0.013 | -0.056 |
| Social | lehren | -0.139 | -0.013 | 0.066 | -0.064 | -0.004 | 0.243 |
| Social | trösten | -0.012 | -0.045 | -0.041 | -0.136 | 0.041 | 0.156 |
| Social | zuhören | -0.079 | -0.049 | -0.052 | 0.114 | -0.062 | -0.039 |
| Enterprising | anwerben | 0.032 | 0.044 | 0.081 | -0.090 | -0.064 | 0.051 |
| Enterprising | auffordern | -0.018 | 0.098 | -0.115 | 0.048 | 0.129 | -0.088 |
| Enterprising | beeinflussen | 0.025 | -0.022 | 0.230 | -0.050 | -0.027 | 0.154 |

| | | | | | | | |
|--------------|---------------|--------|--------|--------|--------|--------|--------|
| Enterprising | durchgreifen | -0.043 | -0.058 | -0.024 | 0.105 | 0.012 | -0.021 |
| Enterprising | durchsetzen | 0.108 | 0.132 | -0.146 | -0.147 | 0.175 | 0.163 |
| Enterprising | erkämpfen | -0.041 | 0.053 | -0.057 | 0.097 | -0.125 | 0.043 |
| Enterprising | fordern | 0.042 | -0.013 | -0.059 | 0.121 | 0.077 | -0.088 |
| Enterprising | führen | -0.002 | -0.031 | -0.102 | 0.578 | -0.013 | 0.073 |
| Enterprising | handeln | 0.022 | -0.110 | 0.123 | -0.148 | -0.087 | 0.092 |
| Enterprising | investieren | 0.017 | 0.025 | 0.161 | -0.067 | -0.015 | 0.014 |
| Enterprising | leiten | -0.058 | -0.041 | -0.002 | -0.227 | 0.156 | 0.094 |
| Enterprising | überreden | 0.000 | -0.036 | -0.033 | -0.210 | -0.009 | 0.068 |
| Enterprising | überzeugen | -0.110 | -0.330 | 0.102 | -0.043 | 0.198 | -0.231 |
| Enterprising | verhandeln | 0.063 | -0.037 | 0.098 | 0.004 | -0.169 | -0.017 |
| Conventional | aufzählen | 0.016 | -0.020 | -0.071 | -0.008 | 0.014 | -0.042 |
| Conventional | aussortieren | 0.015 | -0.060 | -0.087 | -0.033 | 0.099 | -0.052 |
| Conventional | beantragen | 0.235 | -0.013 | -0.110 | -0.040 | -0.214 | -0.029 |
| Conventional | einordnen | 0.168 | -0.094 | 0.086 | -0.053 | 0.018 | 0.075 |
| Conventional | kalkulieren | -0.016 | 0.045 | 0.087 | 0.138 | 0.045 | -0.027 |
| Conventional | kontrollieren | -0.047 | -0.042 | 0.074 | -0.195 | 0.119 | 0.132 |
| Conventional | nummerieren | -0.044 | -0.051 | 0.105 | 0.036 | 0.083 | 0.074 |
| Conventional | ordnen | 0.083 | 0.038 | -0.181 | 0.030 | -0.015 | -0.101 |
| Conventional | überweisen | -0.065 | 0.024 | 0.007 | 0.083 | 0.022 | 0.055 |
| Conventional | verwalten | -0.024 | 0.011 | -0.031 | 0.180 | -0.020 | 0.037 |
| Conventional | zählen | 0.018 | 0.014 | 0.529 | -0.131 | 0.047 | -0.052 |