

# **colloquium:** **new philologies**



*Colloquium: New Philologies* is edited by the Alpen-Adria-Universität Klagenfurt.

**Chief Editor:** Nikola Dobrić

**Co-Chief Editor:** Cristina Beretta

**Section Editors:** Cristina Beretta, Marta Degani, Nikola Dobrić, Angela Fabris,  
Paul Keckeis, René Reinhold Schalleger, Jürgen Struger, Peter Svetina, Giorgio Ziffer

**Technical Editor:** Thomas Hainscho

**Language Support:** Vanessa Erat

#### **World Wide Web**

Visit Colloquium online at <http://colloquium.aau.at>

#### **Legal Information**

*Colloquium* is an open access, peer-reviewed Academic Journal. It is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Colloquium Logo by Gerhard Pilgram; Open Access Logo by Public Library of Science, from Wikimedia Commons (CC BY-SA 3.0); Title page: "0392" by kusakari from [www.openprocessing.org](http://www.openprocessing.org) (CC BY-NC-SA 3.0).

2021 by Alpen-Adria-Universität Klagenfurt

DOI: 10.23963/cnp.2021.6.2

ISSN 2520-3355

# Table of Contents

Colloquium: New Philologies · Vol 6, No 2 (2021)

## Language and Linguistics: Results

- Perspektiven einer Qualitativen Stilometrie am Beispiel stilschichtig  
abgesenkter Lexeme** 1

ELISABETH EDER & ULRIKE KRIEG-HOLZ

- The Role of Errors in Validating a Large-Scale Assessment of Adolescent  
English Writing in Austria** 18

SAMUEL HAFNER & GÜNTHER SIGOTT

- Conceptualisation of Sustainability. A Theoretical and Methodological Frame-  
work for Understanding Metaphors and Narratives of Sustainability** 46

FRANZISCA WEDER & NIKOLA DOBRIĆ

## Literature and Culture: Results

- Mythos versus Utopie. Zur Bedeutung von Kindheit bei Bruno Schulz und  
Theodor W. Adorno** 65

KAROLINE THAIDIGSMANN

# Perspektiven einer Qualitativen Stilometrie am Beispiel stilschichtig abgesenkter Lexeme

**Elisabeth EDER**

Universität Klagenfurt (Austria)

**Ulrike KRIEG-HOLZ**

Universität Klagenfurt (Austria)

## Abstract

Here we outline an approach to stylometry, which intends to be more comprehensive compared to classical stylistic metrics and its commonly used lexical frequency counts. As a prerequisite, such an approach needs language data as a basis for its stylistic analyses. In this paper, we describe the acquisition of two relevant resources: First, we depict collecting and preparing CODE ALLTAG, a German-language email corpus, which contains formal expressions as well as informal and personal in-

teractions, and thus possesses a high stylistic variability. Envisaging the analysis of the vulgar, rough or obscene dimensions of style, we then detail inducing VULGER, a lexical resource covering the lower end of the German language register.

Keywords: *Stylometry, Language resources, Emails, Vulgarity*

(c) Elisabeth Eder & Ulrike Krieg-Holz; elisabeth.eder@aau.at; ulrike.krieg-holz@aau.at

Colloquium: New Philologies, Volume 6, Issue 2 (2021)

doi: 10.23963/cnp.2021.6.2.2

Stable URL: <https://colloquium.aau.at/index.php/Colloquium/article/view/157>

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

## 1 Einleitung

Der linguistische Stilbegriff fokussiert ganz generell die Form der sprachlichen Ausgestaltung von Textstrukturen. Diese Spezifik sprachlicher Formulierungen resultiert aus der Möglichkeit, innerhalb von im Sprachsystem angelegten Varianten auszuwählen. Stil ist deshalb als ein Phänomen der Wahl anzusehen und ein Ergebnis von Entscheidungsprozessen, die sich einerseits an Vorgegebenem, Prototypischem und Musterhaftem orientieren, andererseits immer auch eigenständige Umsetzungen in Verbindung mit individualstilistischen Merkmalen darstellen. Derartige Wahlentscheidungen sind in sämtlichen Kommunikationsformen von größter pragmatischer Relevanz, weil sie das kommunikative Handeln entscheidend prägen und sowohl Textproduktions- als auch Textrezeptionsprozesse erheblich beeinflussen. Denn zum einen können sprachliche Handlungen desselben Typs auf verschiedene Weise durchgeführt werden, zum anderen unterscheiden sich sprachliche Handlungen verschiedenen Typs in stilistischer Hinsicht.

Das Ziel der sprachwissenschaftlichen Stilistik besteht darin, innerhalb von Texten und kommunikativen Zusammenhängen diejenigen Elemente und Strukturen aufzudecken, mit denen das Spezifische der sprachlichen Gestaltung einer kommunikativen Handlung charakterisiert werden kann. Dazu hat sich ein terminologisches Inventar herausgebildet, das geeignet ist, stilistische relevante Einzelelemente – etwa auf den linguistischen Beschreibungsebenen der Lexik oder Grammatik – sowie komplexere stilistische Strukturen (z.B. Stilmittel, Vertextungsstrategien) zu klassifizieren. Dieses stilistische Analyseinventar soll mit Verfahren der automatischen Textklassifikation aus dem Bereich der Computerlinguistik verbunden werden, um (semi-)automatische Stilanalysen im Sinne einer „Qualitativen Stilometrie“ zu ermöglichen.

Zum Standardinventar der klassischen Stilometrie zählen insbesondere lexikalische Häufigkeitsverteilungen und Buchstaben-N-Gramm-Frequenzen. Der Erfolg solcher Verfahren ist nicht zu bestreiten, sie operieren jedoch nur an der Oberfläche des Phänomens Stil als der spezifischen Form sprachlicher Formulierung. Im Folgenden geht es deshalb um die Entwicklung eines sogenannten Stilwerkzeugkastens, der zur automatischen Erkennung und Klassifikation stilistischer Merkmale genutzt werden kann und sich gegenüber bereits existierenden Methoden der klassischen Stilometrie stärker an der Qualität der stilistischen Einzelmerkmale orientiert. Im Zentrum des vorliegenden Beitrags steht dabei insbesondere die Erläuterung der notwendigen infrastrukturellen Voraussetzungen, da Tools dieser Art zunächst nach Sprachressourcen verlangen, auf deren Basis sprachstatistische Berechnungen durchgeführt werden können. Dazu werden zwei einschlägige Ressourcen vorgestellt: ein stilistisch breitbandig angelegtes Textkorpus (CODE ALLTAG) und ein deutschsprachiges Lexikon (VULGER). Letzteres repräsentiert (auf der Basis intersubjektiver Kriterien) eine spezifische Facette sprachlichen Stils,

die Vulgarität im Sinne eines stilschichtig<sup>1</sup> stark abgesenkten Sprachgebrauchs, und soll damit beispielhaft als Fixpunkt für empirisch valide Stilberechnungen fungieren.

Der Beitrag geht zunächst auf die besondere Problematik von E-Mail-Korpora ein und beschreibt den aktuellen Stand der Korpusarbeiten zu CODE ALLTAG (Erhebungsprinzipien, Entwicklungsstand, Annotation und Pseudonymisierung). Im Anschluss daran wird der methodische Ansatz der qualitativen Stilometrie erläutert und am Beispiel des abgesenkten Lexembereichs beschrieben, auf welche Weise dieser durch Verfahren der qualitativen Stilometrie präzisiert und klarer ausdifferenziert werden kann. Dazu wird eine Methode aus dem Bereich der Distributionellen Semantik – die sog. Word Embeddings – vorgestellt und auf deren Eignung für die linguistische Lexik- und Stilbeschreibung eingegangen. Schließlich wird der Aufbau des vulgärsprachlichen Lexikons (VULGER) als ein zentraler Entwicklungsschritt hin zum Stilwerkzeugkasten beschrieben.

## 2 Ein stilistisch breitbandiges Textkorpus: das E-Mail-Korpus CODE ALLTAG

Eine grundlegende Voraussetzung für eine breit angelegte Stilanalyse ist ein hohes Maß an stilistischer Varianz. Die großen deutschsprachigen Textkorpora, wie beispielsweise das DEREKO-Korpus (Kupietz & Lungen 2014) oder das DWDS-Kern- bzw. Zeitungskorpus (Geyken 2007), erfassen primär hoch- bzw. standardsprachlich geprägte Textformen wie Zeitungstexte oder Literatur. Sie spiegeln damit nur einen Teil der gegenwärtigen Erscheinungsformen des Deutschen wider. So lassen sie etwa elektronisch vermittelte Kommunikationsformen und Textsorten, für die spezielle Ausdrucksweisen und Formulierungsmuster typisch sind, ebenso außer Acht wie die verschiedenen Formen von Alltagskommunikation. Dem trägt seit einigen Jahren eine Tendenz in der Korpuslinguistik Rechnung, die sich mit diversen Ausprägungen geschriebener informeller Alltagssprache beschäftigt, wie sie sich in Mikroblog- und Chat-Texten finden (Storrer 2013). Zu nennen ist in diesem Zusammenhang insbesondere DERIK, ein Korpus zur Erfassung computervermittelter Kommunikation in Blogs und Chats (Beißwenger & Lemnitzer 2013). E-Mails berücksichtigt diese Textkollektion jedoch nicht. Diese Lücke schließen Aktivitäten, die am Institut für Germanistik der Universität Leipzig Mitte 2014 begonnen und an der Alpen-Adria-Universität Klagenfurt in Kooperation mit dem Institut für Germa-

---

<sup>1</sup> Die Unterteilung von Stilschichten bzw. Stilebenen ist vor allem in Zusammenhang mit der Kodifikation und Erläuterung des Wortschatzes in Wörterbüchern bekannt. Dabei wird üblicherweise von drei stilistischen Hauptebenen ausgegangen, nämlich „neutral“, „gehoben“ und „abgesenkt“, von denen gerade die letzte weiter unterteilt wird (z.B. „umgangssprachlich“, „derb“, „vulgär“).

nistische Sprachwissenschaft der Friedrich-Schiller-Universität Jena fortgesetzt wurden. Das Ziel dieser Arbeiten bestand im Aufbau eines umfassenden Corpus deutschsprachiger E-Mails (CODE ALLTAG; s.a. Krieg-Holz et al. 2016). Die Kommunikationsform *E-Mail* umfasst zahlreiche Textsorten, die inhaltlich von professionellen bis hin zu ganz persönlichen Interaktionen reichen und somit eine sehr große Bandbreite der performativen Varianz abbilden. Es zeigt sich nahezu das gesamte Kontinuum an Formulierungsmöglichkeiten. Auf der Ebene der Lexik erstreckt sich dies u.a. von stilisch gehoben über neutral und umgangssprachlich bis hin zu expressiven oder vulgären Elementen. Diese im Korpus enthaltene stilistische Varianz soll deshalb als Grundlage für die stilometrischen Untersuchungen dienen.

E-Mail-Korpora sind Korpora, die ausschließlich bzw. überwiegend aus elektronischer Post (E-Mails) bestehen. E-Mails treten selten isoliert auf, sondern sind häufig Teil sog. Threads, also asynchron geführter, thematisch um ein *Subject* (Betreff) zentrierter, konsekutiver Folgen von einzelnen Mails, die den Verlauf von inhaltlichen Austauschen der Schreiber\*innen dokumentieren (Krieg-Holz & Hahn 2016). Somit können solche Korpora dahingehend unterschieden werden, ob sie formal kohärent sind, weil sich Mails über ihre Thread-Struktur explizit aufeinander beziehen, oder inkohärent (ohne Kontextbezug innerhalb des E-Mail-Diskurses). Aufgrund des oft persönlichen Charakters gibt es vergleichsweise wenige solcher E-Mail-Korpora. Zudem sind im juristischen Sinne ausschließlich die Sender\*innen – und eben nicht die Empfänger\*innen – im Besitz des Urheberrechts an ihrer E-Mail (s.a. Krieg-Holz & Hahn 2016). Aus diesem Grund ist es auch für die linguistische Analyse notwendig, zum einen die Zustimmung der E-Mail-Produzent\*innen für die Aufnahme der Texte in ein Korpus zu erlangen, zum anderen ist die Textsammlung in anonymisierter Form zusammenzustellen, um sie nachfolgend für die empirische Untersuchung nutzen zu können. Werden allerdings große Sammlungen von E-Mails von Dritten in einem eigenen, öffentlich zugänglichen Archiv zusammengetragen, die im Besitz keiner eindeutig benennbaren juristischen Person stehen, sind die Rahmenbedingungen weniger restriktiv. Dies ist etwa beim USENET-Newsgroup-Archiv der Fall, das auch innerhalb von CODE ALLTAG als Ressource für ein eigenes Segment genutzt wird (vgl. Krieg-Holz et al. 2016).

Die derzeit verfügbaren E-Mail-Korpora sind überwiegend englischsprachige. Hierzu gehören etwa das ENRON-Korpus (Klimt & Yang 2004), das TREC (Cormack 2007), das W3C-Korpus des World Wide Web Consortiums<sup>2</sup> oder das AUSTRALIAN NATIONAL CORPUS (Lampert 2009). Letzteres enthält ein E-Mail-Segment (Email Australia), das mithilfe eines Aufrufs zur E-Mail-Spende aufgebaut worden ist. Für das Deutsche

---

<sup>2</sup> <https://www.w3.org/>

existierten vor CODE ALLTAG zwei E-Mail-Korpora: ein kleines von Declerck & Klein (1997), das 160 E-Mails umfasst sowie das deutlich größere FLAG-Korpus, das durch die Extraktion von 120.000 Sätzen aus der Internet USENET-Newsgroup generiert wurde (Becker et al. 2003). Angesichts des zugrundeliegenden Forschungsinteresses, der Fehlerannotation, ist das FLAG-Korpus auf sprachliche Verwendungsweisen ausgerichtet, die von standardsprachlicher Schriftlichkeit abweichen (z.B. an Mündlichkeit orientierten Schreibweisen). Demgegenüber hebt sich CODE ALLTAG durch eine größere sprachliche und stilistische Varianz ab (Krieg-Holz & Hahn 2016).

CODE ALLTAG besteht als Gesamtkorpus aus zwei gänzlich verschiedenen Partitionen. Die eine wurde auf der Grundlage des Spenden-Modells gewonnen. Sie wird auch als CODE ALLTAG<sub>S+D</sub> bezeichnet und bildet den Gegenstand der folgenden Erläuterungen. Die andere wird aufgrund ihrer Größenordnung mit CODE ALLTAG<sub>XL</sub> benannt, denn sie setzt sich aus knapp 1,5 Mio. E-Mails zusammen. Diese wurden aus sieben inhaltlich unterschiedlichen Kategorien des deutschsprachigen Teils des Internet USENET-Newsgroup-Mail-Archivs extrahiert und minimal veredelt. Das heißt, es erfolgte die Beseitigung inhaltlicher Verrauschungen durch Fakes und Spams und gegebenenfalls die Auflösung von Zeichenkodierungsproblemen.

CODE ALLTAG<sub>S+D</sub> ist wesentlich kleiner als CODE ALLTAG<sub>XL</sub>. Es besteht derzeit (Stand November 2020) aus knapp 1.500 E-Mails. Jedoch enthält es ergänzend zu den E-Mails demographische Angaben zu den E-Mail-Produzent\*innen.<sup>3</sup> Diese Zusatzinformationen sollten ursprünglich primär für die stilistisch-forensische Analyse genutzt werden, sie sind aber darüber hinaus auch für verschiedenste variations- bzw. soziolinguistische Fragestellungen und Anwendungen nutzbar.

Beim Aufbau des CODE ALLTAG-Korpus wurde ein Ad-hoc-Sample angestrebt, denn eine statistisch repräsentative Auswahl von E-Mails würde nicht nur am Problem der verlässlichen Definition der Grundgesamtheit und am Privatheitsprinzip scheitern, sondern auch an der Zustimmung der Schreiber\*innen zur Aufnahme ihrer E-Mails in die Stichprobe. In einem ersten Schritt wurden Germanistikstudierende in einer Einführungsvorlesung an der Universität Leipzig gebeten, jeweils eine ihrer E-Mails – ohne jegliche Veränderung der Original-E-Mail – in das Korpus zu spenden. Danach wurde über den allgemeinen Verteiler der Universität Leipzig an alle Institute und Verwaltungseinheiten ein Aufruf zur E-Mail-Spende gesendet. Um zu verhindern, dass die Schreiber\*innen allesamt im universitären Umfeld tätig sind, haben wir zudem darum gebeten, auf die Spendenaktion auch im Bekanntenkreis (Familien, Freunde, Vereine usw.) hinzuweisen und so die Zusendung weiterer E-Mails zu initiieren. Die in Leipzig begonnenen Arbei-

---

<sup>3</sup> Für eine detaillierte Beschreibung beider Partitionen des CODE ALLTAG-Korpus vgl. Krieg-Holz et al. (2016).



ten werden seit 2015 am Institut für Germanistik der Alpen-Adria-Universität Klagenfurt fortgesetzt.<sup>4</sup> Hier wurden inzwischen mehrere Initiativen gestartet, um den Zulauf von E-Mails zu steigern. Das Spektrum reichte von der Verteilung von Flyern, über die persönliche Ansprache von Passant\*innen in öffentlichen Räumen und Besucher\*innen sozialer Ereignisse bis hin zur Einbeziehung größerer Firmen und Schulen.<sup>5</sup>

Für die Auswahl der E-Mail wurden keinerlei formale oder inhaltliche Präferenzen vorgegeben. Sie erfolgte nach freien Stücken durch die jeweiligen Spender\*innen. Obwohl das Korpus auch Threads als Bestandteil einzelner E-Mail-Spenden enthält, handelt es sich insgesamt um ein inkohärentes E-Mail-Korpus.<sup>6</sup> Die zur Spende ausgewählte E-Mail wurde jeweils als Weiterleitungs-E-Mail an die Sammeladresse gesendet und dort gespeichert. Mit dem Empfang einer E-Mail wurde dann ein Fragebogen an die jeweiligen Spender\*innen geschickt, mittels dessen demographische Angaben zu den Produzent\*innen der E-Mails erhoben werden sollten. Diese Erhebung umfasste neben konventionellen Sozialdaten (Alter, Geschlecht, Ausbildung, Beruf usw.) vor allem auch Angaben zur Häufigkeit der Nutzung von E-Mails im Alltag, zur regionalen Herkunft und sonstigen Rahmenbedingungen der Autor\*innen in Hinblick auf ihr Sprachverhalten. Besonders relevant ist in diesem Zusammenhang, dass am Ende des Fragebogens eine juristisch eindeutige Zustimmung der Spender\*innen zur Weiterverwendung der zugeschickten E-Mail unter der Voraussetzung, dass deren Identität nicht mehr ersichtlich ist, erbeten wird.<sup>7</sup>

Um eine unrestringierte Weitergabe der E-Mails zu ermöglichen, wurden Wörter und Phrasen, durch die einzelne E-Mail-Verfasser\*innen identifiziert werden könnten, in einem nächsten Schritt folglich pseudonymisiert.<sup>8,9</sup> Im Gegensatz zur Anonymisierung wird sensible, d.h. individuenidentifizierende Information bei der Pseudonymisierung

<sup>4</sup> Da die Kampagne derzeit noch weitergeführt wird, bitten wir nach wie vor – auch die Leser\*innen dieses Beitrags – um E-Mail-Spenden an folgende Adresse: *kodealltag@aau.at*.

<sup>5</sup> Das E-Mail-Segment des AUSTRALIAN NATIONAL CORPUS basiert auf einer anderen Akquisitionsmethode. Dort hatte man anhand eines vordefinierten 8-Kategorienschemas (z.B. Familie, Beschwerden, Liebe/Romantik) einen öffentlichen nationalen Aufruf (Email Australia) veranlasst, der zur Abgabe von über 10.000 E-Mails führte (Lampert 2009).

<sup>6</sup> E-Mail-Korpora gelten dann als formal kohärent, wenn sie sich über ihre Thread-Struktur explizit aufeinander beziehen. Im Falle von CODE ALLTAG werden die Threads in thread-freie einzelne E-Mails aufgespalten und solche, die die Thread-Struktur bewahren.

<sup>7</sup> Dieser rechtlichen Notwendigkeit sind zahlreiche Spender\*innen nicht nachgekommen, so konnte ca. ein Drittel der zugesendeten Datensätze nicht in CODE ALLTAG aufgenommen werden.

<sup>8</sup> Im Falle des AUSTRALIAN NATIONAL CORPUS zeigte die von den Sender\*innen selbst durchgeführte Anonymisierung individueller Daten große Defizite (Lampert 2009), deshalb wurde diese Aufgabe bei CODE ALLTAG nicht delegiert.

<sup>9</sup> Siehe ähnliche Arbeiten zu klinischen Datensätzen, unter anderem Meystre (2015) oder Stubbs et al. (2015a, 2015b, 2017), aber auch allgemeinere Arbeiten von Medlock (2006).

nicht nur durch Platzhalter wie *XXX* ersetzt, was für viele Anwendungsfälle wie auch für den vorliegenden ungeeignet ist, sondern durch realistische Bezeichnungsalternativen. *Irene Adler* könnte beispielsweise durch *Herta Tschach* ersetzt werden. Die Substitution beschränkt sich jedoch nicht nur auf Vor- und Nachnamen (bei Vornamen werden weibliche und männliche getrennt erfasst), sondern inkludiert auch Namen von Organisationen, Firmen oder Institutionen, Usernamen, Datumsangaben, Orts- und Regionsnamen, Straßennamen, Hausnummern, Postleitzahlen, E-Mail-Adressen, URLs und Domains, Telefon- und Faxnummern sowie Passwörter und IDs jederart als individuenidentifizierende Entitäten-Typen. Bevor spezifische textuelle Erwähnungen (sog. entity mentions) dieser Typen jedoch ersetzt werden können, müssen sie im Text erkannt werden. Mithilfe der Annotationssoftware BRAT<sup>10</sup> (Stenetorp et al. 2012) wurden 1.390 (nach damaligem Stand alle) E-Mails von CODE ALLTAG<sub>S+D</sub><sup>11</sup> von drei Annotator\*innen und 1.000 E-Mails CODE ALLTAG<sub>XL</sub> von fünf Annotator\*innen gemäß der verschiedenen Typen manuell annotiert. Diese annotierten E-Mails fungierten als Trainingsdaten für neuronale Machine Learning Modelle zur automatischen Erkennung sensibler Textstellen. Die manuell und maschinell erkannten sensiblen Erwähnungen der oben aufgezählten Entity-Typen wurden im Anschluss daran durch automatisch generierte Alternativen substituiert.<sup>12</sup> Nach der Verschleierung der Identität der E-Mail-Verfasser\*innen ist die pseudonymisierte Version von CODE ALLTAG jetzt für Weiterverwendungen unterschiedlicher Art online zugänglich über <https://github.com/codealltag>.

Diese beiden Partitionen von CODE ALLTAG bilden nun den Ausgangspunkt für die weiteren stilometrischen Untersuchungen.

### 3 Qualitative Stilometrie: Begriffsbestimmung und methodisches Vorgehen

Die Qualitative Stilometrie ist an der Schnittstelle von Linguistik und Computerlinguistik angesiedelt und soll dazu dienen, traditionelle Konzepte aus linguistischen Bereichen wie der Stilistik<sup>13</sup> oder Lexikographie mittels computerlinguistischer Verfahren weiter auszudifferenzieren und zu präzisieren. Dies bildet die Grundlage für die Entwicklung von verschiedenen Werkzeugen, die in Form eines Stilbaukastens für die Analyse digita-

<sup>10</sup> <http://brat.nlplab.org/>

<sup>11</sup> Bei CODE ALLTAG<sub>S+D</sub> wurden zudem Markierungen für unterschiedliche Threads gekennzeichnet, um die Nachrichten später anhand deren zu splitten, sowie Meta-Daten, die anschließend entfernt wurden.

<sup>12</sup> Für eine detaillierte Beschreibung der Erkennungs- und Substitutionsverfahren zur Pseudonymisierung vgl. Eder, Krieg-Holz & Hahn (2019b) und Eder, Krieg-Holz & Hahn (2020).

<sup>13</sup> Vgl. dazu Fleischer, Michel & Starke (1993), Sandig (2006) und Eroms (2008).

ler Editionen und Textkorpora genutzt werden können. Die klassische Stilometrie oder Quantitative Stilistik fokussiert vorrangig Fragestellungen, die die Zuordnung und den Vergleich von Autorenstilen betreffen oder der Identifikation anonymer Autor\*innen dienen. Zum stilometrischen Inventar gehören in Bezug auf solche Anwendungen in der Regel formale Verfahren wie Bestimmung von Satzlängen, Wortlängen und lexikalischen Verteilungen (Eder, Kestemont & Rybicki 2016). Gegenüber derartigen Verfahren ist der hier vorgestellte stilometrische Ansatz insofern stärker qualitativ ausgerichtet, als er zum einen das stilometrische Inventar auf prinzipiell alle stilistisch relevanten sprachlichen Kategorien erweitert. Zum anderen ermöglicht er zu zeigen, wie diese stilistischen Kategorien kodiert sind und wie sie miteinander in Verbindung stehen.

Innerhalb der Stilistik ist es unbestritten, dass die Lexik einen sehr direkten Einfluss auf die stilistische Prägung eines Textes hat. Die Untersuchungen zur Qualitativen Stilometrie setzen deshalb im Bereich der Lexik an und gehen grundsätzlich von der Annahme aus, dass es bestimmte Einschätzungsparameter geben muss, bei denen sich z.B. Unterschiede zwischen einem hochsprachlichem und einem vulgären Text signifikant niederschlagen.<sup>14</sup>

Für die differenzierte Wortschatzbeschreibung erfolgt zunächst eine Orientierung an etablierten stilistischen und lexikographischen Kategorien. Dabei wird auf oberster Ebene zwischen Neutralität und Markiertheit unterschieden. Im Falle von Markiertheit<sup>15</sup> kommen Kategorien wie beispielsweise veraltet, gehoben, fachsprachlich, dialektal/regional, derb oder vulgär zum Tragen (Krieg-Holz & Bülow 2016). Derartige Markierungen werden zunächst einzeln ausdifferenziert und sollen dann jeweils über spezielle Lexika abrufbar sein. Später sollen diese Lexika in den Werkzeugkasten integriert werden und ermöglichen, die linguistische Charakteristik von Texten auszumessen.

Den Ausgangspunkt für die differenzierte, stilistische Klassifikation bildet der Bereich der stilschichtig abgesenkten Lexik. Dieser lässt sich stiltheoretisch zunächst von neutralen Elementen unterscheiden. Von (virtueller) stilistischer Neutralität (Eroms 2008)

<sup>14</sup> Beispiel für vulgären Sprachgebrauch finden sich etwa im Cybermobbing-Korpus von Marx (2017, 186), z.B. „ba alter die geht auf babystrich (isg, pg\_1\_rüd\_stu, 2011-04-14, 12:50:46)“.

<sup>15</sup> Das Konzept der Markiertheit geht ursprünglich auf die Natürlichkeitstheorie zurück, einem linguistischen Ansatz, der die Sprecher\*innen mit ihren produktiven und rezeptiven Fähigkeiten in den Vordergrund stellt. Anwendung fand dies vor allem in der Phonologie und Morphologie (vgl. Mayerthaler 1981, Wurzel 1984), wobei es im Wesentlichen darum ging, was für den Sprecher besser oder schlechter ist, was weniger oder mehr markiert ist. Eroms (2008, 60f.) überträgt die Differenzierung zwischen Markiertheit und Neutralität auf die Stilistik und unterscheidet zwei Faktoren, die zu virtueller stilistischer Neutralität führen können: Systemzwang und Systemneutralisierung. Ersteres liegt im Falle eines einelementigen Paradigmas vor, in dem es keine Varianten gibt, also überhaupt nur ein Ausdruck zur Verfügung steht. Um Systemneutralisierung handelt es sich, wenn innerhalb eines Wortfeldes (im Sinne einer Synonymengruppe) ein Ausdruck ein geringeres stilistisches Potential hat, weil er in allen Kommunikationsbereichen angemessen ist.

wird dann gesprochen, wenn ein Lexem ohne Einschränkung in allen Kommunikationsbereichen und Textsorten vorkommen kann ohne eine besondere Wirkung zu erzielen. Auch innerhalb der Lexikographie geht man i.d.R. von einem unmarkierten Zentrum und einer markierten Peripherie aus. Dementsprechend werden Lemmata mit speziellen Markierungen bzw. diasystematischen Angaben versehen. Eine Bedeutung für den abgesenkten Bereich haben dabei ganz verschiedene Kriterien oder Mikrosysteme. Hierzu gehören die Markierungsarten *diamedial* (z.B. umgangssprachlich), *diastatisch* (z.B. Slang), *diaphasisch* (z.B. informell), *diatopisch* (z.B. bairisch) und *diaevaluativ* (z.B. pejorativ). Sie werden nicht durchgehend übereinstimmend angewendet, sondern gehen vielfach ineinander über und wirken unscharf. In der aktuellen Diskussion um Cyber-Mobbing und Hassrede kommen weitere Benennungen dazu, die versuchen, diverse Abstufungen von Beleidigungen, Beschimpfungen abzubilden. Sie fallen primär in den diaevaluativen Bereich. Eine Skala, die in diesem Zusammenhang in computerlinguistischen Untersuchungen angewendet wurde, unterscheidet zwischen *profanity* – *insult* – *abuse* (Wiegand et al. 2018).

Word Embeddings etablieren sich zunehmend als neuartige korpuslinguistische Methode auf der Grundlage der distributionellen Semantik. Sie werden im Folgenden für die automatische Stilbeschreibung genutzt, wobei sie zum einen dazu dienen sollen, die stilistische Qualität von Texten auszumessen, und so etwa in Bezug auf Dichotomien wie *formell* – *informell*, *vulgärsprachig* – *hochsprachig* valide Aussagen darüber zu ermöglichen, worin sich die stilistische Gestaltung dieser Texte unterscheidet. Somit könnten auch Graduierungen vorgenommen werden (sehr vulgär, weniger vulgär usw.). Zum anderen verwenden wir Word Embeddings für den Lexikonaufbau. Mit Hilfe von Word Embeddings können die lexikalisch-semantische Ähnlichkeiten zwischen Lexemen berechnet werden. Grundlage hierfür sind Vektorrepräsentationen, die sich aus dem Kontext von Wörtern in Sätzen ableiten. Damit können u.a. Daten zu kookkurrierenden Elementen generiert und Analogieschlüsse gezogen werden (Mikolov et al. 2013).

Um die Subjektivität der Annotationen (z.B. Abstufungen zwischen vulgärsprachlich und standardsprachlich) zu verringern, integriert die Methode der Qualitativen Stilometrie Crowdsourcing-Verfahren zur Einschätzung semantischer Unterscheidungen. Diese nutzen eine möglichst heterogene Menge von unbekannten Akteur\*innen, um Aufgaben, die traditionell intern von einigen wenigen Personen durchgeführt werden, zu übernehmen und auf eine breitere quantitative Basis zu stellen. Über das Internet und diverse Crowdsourcing-Anbieter lässt sich durch derartige Auslagerungsstrategien die Zahl an Einschätzer\*innen beträchtlich erhöhen.

## 4 Lexikon für stilistisch abgesenkte Sprache: VULGER

Eine zentrale Grundlage für den Stilwerkzeugkasten sollen Lexika bilden, in denen bestimmte Bereiche stilistisch relevanter Markierungen so präzise wie möglich abgebildet werden. In einem ersten Schritt wurde dazu zunächst ein Lexikon für den abgesenkten Bereich aufgebaut (VULGER, vgl. Eder, Krieg-Holz & Hahn (2019a)). Der Prozess zur Erstellung orientiert sich an der Arbeit von Wiegand et al. (2018), wobei dieses Lexikon anders als herkömmliche Ansätze, die Lexeme in Kategorien (zum Beispiel binär abgesenkt versus nicht abgesenkt oder abwertend, derb, vulgär, etc.) einteilen, lexikalische Elemente auf einer Skala von neutral bis vulgär graduieren soll. Zudem erfolgte die Annotation der Grade nicht durch einzelne Lexikograph\*innen, sondern wurde mittels Crowdsourcing durchgeführt.

### 4.1 Aufbau

Zuerst wurde für den Aufbau des Lexikons auf schon vorhandene lexikalische Ressourcen zurückgegriffen. Aus dem deutschen WIKTIONARY und dem deutschen Online-Wörterbuch OPENTHESAUROS wurden alle Einträge, die als vulgär, derb oder abwertend<sup>16</sup> klassifiziert waren, extrahiert, wobei vorerst lediglich Einzellexeme Berücksichtigung fanden, da klassische Word Embeddings einzelwortbasiert sind. Vorhandene Wortbildungselemente mit klarem Bezug zu abgesenkter Sprache, wie *geil*, *scheiß*, *drecks*, deren Vulgarität je nach Komposition variieren kann, wurden dementsprechend durch Komposita mit diesen Elementen, die in den im Folgenden beschriebenen Korpora beziehungsweise Word Embeddings enthalten waren, ersetzt. Die daraus resultierende Liste abgesenkter Lexeme wurde um distributionell ähnliche Wörter erweitert (siehe auch Tulkens et al. (2016), Wiegand et al. (2018)). Dazu wurden zum einen FASTTEXT-Word Embeddings (Grave et al. 2018) verwendet, zum anderen wurden aus CODE ALLTAG und dem DORTMUNDER CHAT KORPUS (Beißwenger 2013) WORD2VEC-Embeddings (Mikolov et al. 2013) erzeugt. Für letzteres wurde das GENSIM-Modul (Řehůřek & Sojka 2010) genutzt, das auch zur Berechnung der semantisch benachbarten Wörter auf Basis der genannten Word Embeddings eingesetzt wurde. Die mit diesem Vorgehen einhergehenden flektierten Formen und Falschschreibungen wurden manuell bearbeitet und gegebenenfalls beseitigt. Hingegen wurden berechnete semantische Nachbarwörter, die mutmaßlich neutral sind, nicht aus der Liste entfernt, weil sie für unseren Anwendungsfall der Graduierung von Vulgarität von Nutzen sind.

Insgesamt umfasst das in diesem Schritt entstandene Basislexikon für die weitere Annotation 3.300 Einträge.

<sup>16</sup> Inklusive der Abkürzungen vulg., vul. und abw.

## 4.2 Annotation

Die Annotation des Lexikons wurde neben der Crowdsourcing-Plattform FIGURE EIGHT<sup>17</sup> vor allem mit dem deutschen Crowdsourcing-Anbieter CLICKWORKER<sup>18</sup> umgesetzt. Für die Skalierung der Einträge selbst kam Best-Worst-Scaling (BWS) zum Einsatz. Diese von Louviere, Flynn & Marley (2015) entwickelte Methode verspricht Annotationen von hoher Qualität, wie Kiritchenko & Mohammad (2017) zeigen, die BWS zur Bewertung emotionaler Sprache eingesetzt haben (Kiritschenko & Mohammad 2016, 2017). Beim BWS werden den Annotator\*innen jeweils  $n$  Begriffe vorgelegt, wobei  $n$  üblicherweise 4 ist. Dann müssen sie aus diesem  $n$ -Tupel den besten Begriff, der das gewählte Kriterium am besten erfüllt, und den schlechtesten Begriff, auf den die Eigenschaft folglich am wenigsten zutrifft, wählen. In dem hier beschriebenen konkreten Fall wurde dementsprechend nach dem neutralsten und nach dem vulgärsten Wort in einem 4er-Tupel gefragt. Dabei wurden die Einschätzer\*innen mit folgendem Text an ihre Aufgabe herangeführt:

---

Wörter können neutral sein, das heißt sie können überall (Nachrichten etc.) verwendet werden, z.B. *Schmetterling*. Sie können aber auch auf die Umgangssprache beschränkt (z.B. *kotzen*) oder derb (z.B. *verrecken*) und schon richtig vulgär (z.B. *Pissfotze*) sein. Dabei sind die Übergänge fließend. In diesem Job sollen Sie bestimmen, wie vulgär Wörter sind.

Ihnen werden vier Begriffe präsentiert. Zu diesen bekommen Sie jeweils zwei Aufgaben beziehungsweise Fragen:

1. Neutralstes Wort

Erstens sollen Sie den Begriff wählen, der von diesen vier Wörtern am neutralsten ist, also auf einer Skala von neutral über umgangssprachlich, abwertend und derb bis hin zu vulgär am ehesten in Richtung neutral geht.

2. Vulgärstes Wort

Zweitens sollen Sie den Begriff wählen, der von diesen vier Wörtern am vulgärsten ist, also auf der Skala am nächsten bei vulgär zu finden ist.

---

<sup>17</sup> <https://www.figure-eight.com/>

<sup>18</sup> <https://www.clickworker.de/>

Eine Beispielaufgabe, wie sie bei CLICKWORKER aussehen könnte, zeigt folgendes Bild.

### FOKUSWÖRTER

**Wort 1:** Joghurtbecher

**Wort 2:** saugeil

**Wort 3:** Scheißregierung

**Wort 4:** Hackfresse

**Bitte wählen Sie das NEUTRALSTE Wort.\***

- ☐ Wort 1
- ☐ Wort 2
- ☐ Wort 3
- ☐ Wort 4

**Bitte wählen Sie das VULGÄRSTE Wort.\***

- ☐ Wort 1
- ☐ Wort 2
- ☐ Wort 3
- ☐ Wort 4

Abbildung 1: Beispielaufgabe mit 4er-Tupel auf der Crowdsourcing-Plattform CLICKWORKER

Um aus den 3.300 Einträgen im Basislexikon 6.600 4er-Tupel zu generieren, wurde das BWS-Tool<sup>19</sup> von Kiritchenko & Mohammad (2016, 2017) verwendet, das die Wort-Tupel willkürlich zusammenstellt unter der Voraussetzung, dass eine Tupelzusammensetzung nur einmal vorkommt und ein Wort nicht mehr als einmal in acht verschiedenen Tupeln auftritt. Jedes der erzeugten Tupel wurde fünf Mal von verschiedenen Crowdworkern bearbeitet. Insgesamt wurden die 3.300 Basiswörter folglich mit 33.000 Bewertungen versehen. Ebenfalls auf ein Programm von Kiritchenko & Mohammad (2016, 2017) zurückgreifend wurde anschließend Counts Analysis (Orme 2009) angewendet, um aus der Gesamtheit der Bewertungen einzelne Scores zu berechnen. Dabei wird der prozentuale Anteil von Bewertungen, in denen ein Wort als schlechtestes (d. h. vulgärstes) gewählt wurde, von dem der Bewertungen als bestes Wort subtrahiert, was Scores zwischen +1 (am neutralsten) und -1 (am vulgärsten) ergibt.

<sup>19</sup> <http://www.saifmohammad.com/WebPages/BestWorst.html>

### 4.3 Erweiterung

An Arbeiten zur Emotionserkennung von Lexemen (Li et al. 2017, Buechel & Hahn 2018) orientiert, wurde das Lexikon in einem nächsten Schritt automatisch erweitert, indem mithilfe von Verfahren des Maschinellen Lernens Scores für noch unberücksichtigte, neue Wörter berechnet wurden. Das Basislexikon fungierte dabei als Trainingsgrundlage für ein Ridge Regression-Modell.<sup>20</sup> Dieses lernt aus Word Embeddings, die als Features für die einzelnen Einträge im Lexikon dienten, die passenden Scores.<sup>21</sup> Als Word Embeddings wurden FASTTEXT-Embeddings (Grave et al. 2018) verwendet, wobei dort nicht vorhandene Wörter unberücksichtigt blieben. Nach dem Training wurde das Modell auf neue Wörter angewendet, die thematisch in Frage kommen. Schimpfwörter weisen offenkundig Überschneidungen mit abgesenkten Lexemen auf. Zwar sind nicht alle vulgären Wörter auch Schimpfwörter (bspw. *pissen*) und genauso haben nicht alle Beschimpfungen denselben Grad an Vulgarität (bspw. *Blödmann*, *Jasager* oder *Clown*), dennoch hat der hier verwendete Ansatz der Skalierung genau hier seine Stärke und kann den unterschiedlichen Vulgaritätsgraden gerecht werden.

Zunächst wurden Scores für drei deutsche Schimpfwortlisten errechnet: HYPERHERO<sup>22</sup>, INSULT.WIKI<sup>23</sup> und SCHIMPFWOERTER.DE<sup>24</sup>, wobei Lexeme, die im Basislexikon schon vorhanden waren, ausgeklammert wurden. Gleiches gilt für Einträge, für die auch ohne deren Groß- und Kleinschreibung zu beachten kein entsprechendes FASTTEXT-Word Embedding gefunden wurde, wodurch sehr seltene Wörter ebenfalls exkludiert werden konnten. Auf diesem Weg konnte das Lexikon um 2.046 Wörter erweitert werden.

Außerdem wurde Gebrauch von Korpora gemacht, die als Trainingsmaterial für die Erkennung von Hasssprache in deutschen Tweets dienen, da angenommen werden kann, dass auch sie tendenziell einen abgesenkteren Sprachgebrauch aufweisen. IWG HATE-SPEECH<sup>25</sup> (Ross et al. 2016) enthält ungefähr 500 deutsche Tweets, während das Korpus zur GERM EVAL 2018 Challenge<sup>26</sup> (Wiegand et al. 2018b) mit mehr als 8.500 Tweets wesentlich größer ist. Die Tweets beider Korpora wurden von menschlichen Annotator\*innen neben einer differenzierteren Klassifizierung binär als Hasssprache und keine

<sup>20</sup> Dazu wurde die Implementierung von SKLEARN.ORG mit den Standardparametern verwendet.

<sup>21</sup> Weitere Informationen und eine Evaluation zu unterschiedlichen Regressionsmodellen und Word Embeddings finden sich in Eder, Krieg-Holz & Hahn (2019a).

<sup>22</sup> <http://www.hyperhero.com/de/insults.htm>

<sup>23</sup> <http://www.insult.wiki/wiki/Schimpfwort-Liste>

<sup>24</sup> <https://www.schimpfwoerter.de/>

<sup>25</sup> [https://github.com/UCSM-DUE/IWG\\_hatespeech\\_public](https://github.com/UCSM-DUE/IWG_hatespeech_public)

<sup>26</sup> <https://projects.cai.fbi.h-da.de/iggsa/>



Hasssprache bewertet.<sup>27</sup> Aus allen Tweets, die mindestens einmal als Hasssprache kategorisiert wurden, wurden Wörter extrahiert, wobei Stoppwörter, Hashtags und Token, die kürzer als 4 Buchstaben sind oder nicht-alphabetischen Zeichen mit Ausnahme von Bindestrichen enthalten, keine Berücksichtigung fanden. Die ausgelesenen Wörter wurden anschließend mit SPACY<sup>28</sup> (Honnibal & Montani 2017) weitestmöglich lemmatisiert und bezüglich Groß- und Kleinschreibung normalisiert. Auch hier wurden Einträge ohne entsprechendes Word Embedding exkludiert, um nicht nur seltene, sondern gegebenenfalls auch orthographisch inkorrekte Token zu entfernen. Trotz dieser Bereinigungsstrategien ist die auf diese Art und Weise akquirierte Wortliste mit 5.700 Einträgen stärker verrauscht als die Lexeme, die aus lexikalischen Ressourcen wie Schimpfwortlisten extrahiert wurden.

Mit 3.300 Lexemen aus dem Basislexikon, 2.046 Schimpfwörtern und den 5.700 Wörtern aus Tweets mit Hasssprache umfasst das finale erweiterte VULGER-Lexikon nun insgesamt 11.046 Einträge.

Mit dem vorliegenden Beitrag sollte gezeigt werden, auf welche Weise das vulgärsprachliche Lexikon VULGER aufgebaut wurde. Es stellt einen wichtigen Schritt für die Entwicklung eines Stilwerkzeugkastens dar, der zur automatischen Erkennung und Klassifikation stilistischer Merkmale genutzt werden soll und sich gegenüber bereits existierenden Methoden der klassischen Stilometrie stärker an der Qualität der stilistischen Einzelmerkmale orientiert, um (semi-)automatische Stilanalysen im Sinne einer „Qualitativen Stilometrie“ zu ermöglichen. Im Mittelpunkt der Ausführungen stand die Erläuterung der notwendigen infrastrukturellen Voraussetzungen, wie auch der Akquise eines stilistisch breitbandig angelegten E-Mail-Korpus. Zunächst auf Basis von CODE ALLTAG sollen mithilfe des Stilbaukastens und VULGER Analysen dieses stilistischen Variantenreichtums ermöglicht werden.

## Literatur

- Becker, Markus, Andrew Bredenkamp, Berthold Crysmann und Judith Klein. 2003. „Annotation of error types for German newsgroup corpus.“ In *Treebanks. Building and Using Parsed Corpora*, hrsg. v. Anne Abeillé, 89–100. Dordrecht: Springer Netherlands.
- Beißwenger, Michael und Lothar Lemnitzer. 2013. „Aufbau eines Referenzkorpus zur deutschsprachigen internetbasierten Kommunikation als Zusatzkomponente für die Korpora im Projekt ‚Digitales Wörterbuch der deutschen Sprache‘ (DWDS)“. In *Journal for Language Technology and Computational Linguistics* 28 (2): 1–22.

<sup>27</sup> Für IWG HATESPEECH: HatespeechOrNot? Yes oder No; für GERMEVAL 2018: Offense oder Other.

<sup>28</sup> <https://spacy.io/>

- Beißwenger, Michael. 2013. „Das Dortmunder ChatKorpus.“ In *Zeitschrift für germanistische Linguistik*, 41(1): 161–164.
- Buechel, Sven und Udo Hahn. 2018. „Word emotion induction for multiple languages as a deep multi-task learning problem.“ In *NAACL-HLT 2018 — Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans, Louisiana, USA, June 1-6, 2018, volume 1: Long Papers*, 1907–1918. Stroudsburg, PA: Association for Computational Linguistics (ACL).
- Cormack, Gordon V. 2007. „TREC 2007 Spam Track overview.“ In *TREC 2007 – Proceedings of the 15th Text REtrieval Conference. Gaithersburg, Maryland, November 5–9, 2007*.
- Declerck, Thierry und Judith Klein. 1997. „Ein Email-Korpus zur Entwicklung und Evaluierung der Analysekomponente eines Terminvereinbarungssystems.“ In *DGfS-CL '97 – Proceedings der 6. Fachtagung der Sektion Computerlinguistik der Deutschen Gesellschaft für Sprachwissenschaft: Integrative Ansätze in der Computerlinguistik. Heidelberg, Deutschland, 8.–10. Oktober 1997*.
- Eder, Elisabeth, Ulrike Krieg-Holz und Udo Hahn. 2019a. „At the Lower End of Language – Exploring the Vulgar and Obscene Side of German.“ In *Proceedings of the Third Workshop on Abusive Language Online. Florence, Italy, August 1, 2019*, 119–128. Stroudsburg, PA: Association for Computational Linguistics (ACL).
- Eder, Elisabeth, Ulrike Krieg-Holz und Udo Hahn. 2019b. „De-Identification of Emails: Pseudonymizing Privacy-Sensitive Data in a German Email Corpus.“ In *Proceedings of Recent Advances in Natural Language Processing (RANLP). Varna, Bulgaria, 2–4 September 2019*, 259–269.
- Eder, Elisabeth, Ulrike Krieg-Holz und Udo Hahn. 2020. „CodE Alltag 2.0 – A Pseudonymized German-Language Email Corpus.“ In *Proceedings of the International Conference on Language Resources and Evaluation (LREC). Marseille, France, 13–15 May 2020*, 4468–4479.
- Eder, Maciej, Mike Kestemont und Jan Rybicki. 2016. „Stylometry with R: a package for computational text analysis.“ In *R Journal*, 16 (1):107–121.
- Eroms, Hans-Werner. 2008. *Stil und Stilistik. Eine Einführung*. Berlin: Erich Schmidt.
- Fleischer, Wolfgang, Michel, Georg und Starke, Günter. 1993. *Stilistik der deutschen Gegenwartssprache*. Frankfurt a. M. u.a.: Lang.
- Geyken, Alexander. 2007. „The DWDS corpus: a reference corpus for the German language of the 20th century.“ In *Collocations and Idioms: Corpus-based Linguistic and Lexicographic Studies*, hrsg. v.Christiane Fellbaum, 23–40. London: Continuum.
- Grave, Edouard, Piotr Bojanowski, Prakhar Gupta, Armand Joulin und Tomáš Mikolov. 2018. „Learning word vectors for 157 languages.“ In *LREC 2018 — Proceedings of the 11th International Conference on Language Resources and Evaluation. Miyazaki, Japan, May 7-12, 2018*, 3483–3487. Paris: European Language Resources Association (ELRA).
- Kiritchenko, Svetlana und Saif M. Mohammad. 2016. „Capturing reliable fine-grained sentiment associations by crowdsourcing and best-worst scaling.“ In *NAACL-HLT 2016 — Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego, California, USA, June 12-17, 2016*, 811–817. Stroudsburg, PA: Association for Computational Linguistics (ACL).

- Kiritchenko, Svetlana und Saif M. Mohammad. 2017. „Best-worst scaling more reliable than rating scales: a case study on sentiment intensity annotation.“ In *ACL 2017 — Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, British Columbia, Canada, July 30-August 4, 2017, volume 2: Short Papers*, 465–470. Stroudsburg, PA: Association for Computational Linguistics (ACL).
- Klimt, Bryan und Yiming Yang. 2004. „The Enron corpus: a new dataset for email classification research.“ In *ECML 2004 – Proceedings of the 15th European Conference on Machine Learning. Pisa, Italien, 20.–24. September 2004*, 217–226 (LNCS, 3201). Berlin, Heidelberg: Springer.
- Krieg-Holz, Ulrike, Christian Schuschnig, Franz Matthies, Benjamin Redling und Udo Hahn. 2016. „CODE ALLTAG: A German-language e-mail corpus.“ In *LREC 2016 — Proceedings of the 10th International Conference on Language Resources and Evaluation. Portorož, Slovenia, 23-28 May 2016*, 2543–2550. Paris: European Language Resources Association (ELRA-ELDA).
- Krieg-Holz, Ulrike und Lars Bülow. 2016. *Linguistische Stil- und Textanalyse. Eine Einführung*. Tübingen: Narr/Francke/Attempto.
- Krieg-Holz, Ulrike und Udo Hahn. 2016. „CODE ALLTAG: Ein deutsches E-Mail-Korpus für die Forensische Linguistik.“ In *Performativität in Sprache und Recht*, hrsg. v. Lars Bülow, Jochen Bung, Rüdiger Harnisch und Rainer Wernsmann, 245–264. Berlin, Boston: de Gruyter.
- Kupietz, Marc und Harald Lungen. 2014. „Recent developments in DeReKo.“ In *LREC 2014 – Proceedings of the 9th International Conference on Language Resources and Evaluation. Reykjavik, Island, 26.–31. Mai 2014*, 2378–2385.
- Lampert, Andrew. 2009. „Email in the Australian National Corpus.“ In *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus: Mustering Languages*. 55–60. Somerville, MA: Cascadilla Proceedings Project.
- Louviere, Jordan J., Terry N. Flynn, and A. A. J. Marley. 2015. *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge: Cambridge University Press.
- Marx, Konstanze. 2017. *Diskursphänomen Cybermobbing. Ein internetlinguistischer Zugang zu [digitaler] Gewalt*. (Diskursmuster – Discourse Patterns 17). Berlin, Boston: De Gruyter.
- Mayerthaler, Willy. 1981. *Morphologische Natürlichkeit*. Wiesbaden: Akademische Verlagsgesellschaft Athenaion.
- Medlock, Ben. 2006. „An introduction to NLP-based textual anonymisation.“ In *LREC 2006 – Proceedings of the 5th International Conference on Language Resources and Evaluation. Genua, Italien, 22.–28. Mai 2006*, 1051–1056.
- Meystre, Stéphane M. 2015. „De-identification of unstructured clinical data for patient privacy protection.“ In *Medical Data Privacy Handbook*, hrsg. v. Aris Gkoulalas-Divanis und Grigorios Loukides, 697–716. Cham, Heidelberg, New York, Dordrecht, London: Springer International Publishing.
- Mikolov, Tomáš, Ilya Sutskever, Kai Chen, Gregory S. Corrado und Jeffrey Dean. 2013. „Distributed representations of words and phrases and their compositionality.“ In *Advances in Neural Information Processing Systems 26 — NIPS 2013. Proceedings of the 27th Annual Conference on Neural Information Processing Systems. Lake Tahoe, Nevada, USA, December 5-10, 2013*, 3111–3119. Red Hook, NY: Curran Associates, Inc.

- Orme, Bryan. 2009. „Maxdiff analysis: simple counting, individual-level logit, and HB.“ *Sawtooth Software, Inc.*
- Řehůřek, Radim und Petr Sojka. 2010. „Software framework for topic modelling with large corpora.“ In *Proceedings of the Workshop on New Challenges for NLP Frameworks @ LREC 2010, La Valletta, Malta, May 22, 2010*, 45–50. Paris: European Language Resources Association (ELRA).
- Sandig, Barbara. 2006. *Textstilistik des Deutschen*. Berlin, New York: de Gruyter.
- Stenetorp, Pontus, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou und Jun'ichi Tsujii. 2012. „BRAT: a Web-based tool for NLP-assisted text annotation.“ In *EACL 2012 — Proc. of the 13th Conf. of the European Chapter of the Association for Computational Linguistics: Demonstrations, Avignon, France, April 25-26, 2012*, 102–107.
- Storrer, Angelika. 2013. „Sprachstil und Sprachvariation in sozialen Netzwerken.“ In *Die Dynamik sozialer und sprachlicher Netzwerke. Konzepte, Methoden und empirische Untersuchungen an Beispielen des WWW*, hrsg. v. Barbara Frank-Job, Alexander Mehler und Tilmann Sutter, 329–364. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stubbs, Amber und Özlem Uzuner. 2015a. „Annotating longitudinal clinical narratives for de-identification: the 2014 i2b2/UTHealth corpus.“ In *Journal of Biomedical Informatics* 58 (Supplement): 20–29.
- Stubbs, Amber, Özlem Uzuner, Christopher Kotfila, Ira, Goldstein und Peter Szolovits. 2015b. „Challenges in synthesizing surrogate PHI in narrative EMRs.“ In *Medical Data Privacy Handbook, hrsg. v. Aris Gkoulalas-Divanis und Grigorios Loukides*, 717–735. Cham, Heidelberg, New York, Dordrecht, London: Springer International Publishing.
- Stubbs, Amber, Michele Filannino und Özlem Uzuner. 2017. „De-identification of psychiatric intake records: overview of 2016 CEGS NGRID Shared Tasks Track 1.“ In *Journal of Biomedical Informatics* 75 (Supplement): 4–18.
- Tulkens, Stéphan, Lisa Hilde, Elise Lodewyckx, Ben Verhoeven und Walter Daelemans. 2016. „A dictionary-based approach to racism detection in Dutch social media.“ In *TA-COS 2016 — Proceedings of the Workshop on Text Analytics for Cybersecurity and Online Safety @ LREC 2016, Portorož, Slovenia, 23 May 2016*, 11–17.
- Wiegand, Michael, Josef Ruppenhofer, Anna Schmidt und Clayton Greenberg. 2018. „Inducing a lexicon of abusive words: a feature-based approach.“ In *NAACL-HLT 2018 — Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, Louisiana, USA, June 1-6, 2018, volume 1: Long Papers*, 1046–1056. Stroudsburg, PA: Association for Computational Linguistics (ACL).
- Wurzel, Wolfgang U. 1984. *Flexionsmorphologie und Natürlichkeit*. Studia Grammatica XXI. Berlin: Akademie-Verlag.

# The Role of Errors in Validating a Large-Scale Assessment of Adolescent English Writing in Austria

Samuel HAFNER

University of Klagenfurt (Austria)

Günther SIGOTT

University of Klagenfurt (Austria)

## Abstract

This study investigates errors in a sample of 50 written performances of Austrian learners of English collected in the 2009 baseline study for the Austrian Educational Standards-Based Writing Test for English at grade 8 (E8 Standards Writing Test). The research aims to contribute to the validation of this large-scale assessment by studying the relationship between errors (described using the *Scope – Substance* error taxonomy) and human ratings awarded to writing performances. The results add to the validity evidence of the E8 Standards Writing Test. There is a negative relationship between human ratings and the presence of errors; a low error density is associated with higher ratings and a high error density with lower ratings. *Substance* WORD, CLAUSE, and TEXT error densities play an important role in the rating in most dimensions; errors with a larger *scope* also have a

strong effect. By highlighting aspects of errors to which raters seem to be sensitive, these findings constitute evidence of context validity. At the same time, the findings are relevant to theory-based validity by concretising areas of competence that learners need to develop in order to receive higher ratings. While errors are important determinants of the ratings, additional factors, presumably positive features, must be at play as the accuracy of the regression models is low to moderate. This should in fact be the case since the E8 rating scale refers to negative as well as positive features.

Keywords: *Scope – Substance error taxonomy, E8 Standards Writing Test, Austrian Educational Standards in English, Error analysis, Validation, Validity, Context validity; Theory-based validity*

(c) Samuel Hafner & Günther Sigott; samuel.hafner@aau.at; guenther.sigott@aau.at

Colloquium: New Philologies, Volume 6, Issue 2 (2021)

doi: 10.23963/cnp.2021.6.2.4

Stable URL: <https://colloquium.aau.at/index.php/Colloquium/article/view/143>

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

## 1 Introduction

Analytic rating scales for writing, and the rating scale used in the Austrian Educational Standards-Based Writing Test for English at grade 8 (E8 Standards Writing Test) in particular, describe levels of performance on each of several dimensions by means of so-called performance level descriptors (e.g. Cizek and Bunch 2007, 46). These are supposed to guide trained raters in the process of judging the writing performances with the aim of maximising intra- and interrater agreement, which, in turn, is considered as an indication of validity. In fact, research into the validity of ratings tends to rely on rater agreement as an indication of validity. However, there is little evidence on the extent to which errors, which are only one aspect that is covered by the performance level descriptors in the rating scales, play a role in the raters' rating behaviour. More particularly, little is known about the extent to which the raters actually pay attention to errors, and about the kinds of error to which they are sensitive, when forming their judgments.

This study aims to shed light on this aspect of the rating process in the E8 Standards Writing Test. More specifically, this research aims to contribute to the validation of this test by investigating the role that errors play in the rating of writing performances using an analytic rating scale. The study identifies the most common errors in the writing of Austrian learners of English at the age of around 14 years and, in particular, focuses on the relationship between errors and human ratings awarded to writing performances generated in the context of the E8 Standards Writing Test. The errors in the writing performances are identified and categorised by means of the *Scope – Substance* error taxonomy (Dobrić and Sigott 2014).

In order to address the role of errors in the rating of the E8 Standards Writing Test, the following research questions are investigated:

- RQ1. In terms of errors described using the *Scope – Substance* error taxonomy, what are the most common errors in texts written by Austrian learners of English at grade 8?
- RQ2. What is the relationship between errors described using the *Scope – Substance* error taxonomy and the human ratings in the four dimensions of *Grammar*, *Vocabulary*, *Coherence & Cohesion*, and *Task Achievement*?

## 2 Theoretical Framework

### 2.1 Relation to Validation Theory

Studying the influence of errors on ratings of writing performances addresses aspects of validity. More concretely, this study investigates the relationship between the error incidence and the ratings which were awarded by trained raters using an analytic rating

scale. By doing this, it aims to identify error types that are typical of performances with a lower rating. Avoiding these error types constitutes a challenge which test takers have to master. Avoiding these errors thus constitutes aspects of task difficulty. Identifying the kinds of errors that constitute such difficulty contributes to identifying the construct underlying the E8 Standards Writing Test. In fact, understanding difficulty-generating features in any test is central to understanding what the test tests. In the assessment of writing, difficulty-generating features stem from the test context, such as the task set, the rating criteria, and the raters. The rating criteria are an important source of difficulty-generating features. However, raters looking at the rating criteria per se is no guarantee that they will follow these criteria to the letter, either because they interpret them in their own way or because they develop their own rating strategies when they encounter problems in applying the criteria (Lumley 2005). Therefore, evidence is needed of what the raters actually do when they rate a performance. This evidence can be collected by trying to access the raters' cognitive processes during rating or by observing features of performances that tend to cooccur with individual levels awarded to these performances by the raters. These cooccurrences, or correlations, are empirical evidence which suggests what it is in the performances that raters actually pay attention to in the rating process. These correlations can, however, only be interpreted as suggestive evidence since they are statistical, not causal, relationships. Nevertheless, they indicate features which cooccur with particular ratings. It follows that changes in the occurrence of these features will go hand in hand with changes in the rating. Features correlated with low ratings will, when avoided, make higher ratings likely. The necessity of avoiding such errors, then, can be considered to be a difficulty-generating feature. Consequently, identifying such errors contributes to our understanding of what the test tests, or, put another way, to our ability to identify the test construct. This is why Sigott (2004, 51) has referred to this aspect of validation as *construct identification*.

In terms of Weir's (2005) validation framework, the study addresses *context validity* and *theory-based validity*. Context validity refers to the totality of contextual factors that influence the production of test-taker performance as well as to the criteria for correctness that are applied in scoring the performance. Raters' sensitivity to error types constitutes such expectations of correctness. Identifying the errors that raters attend to when rating therefore provides evidence of context validity. This also contributes to our understanding of rater cognition (Dobrić 2020). From the writer's perspective, avoiding such errors constitutes challenges, or sources of difficulty (Dobrić et al. 2021). Specifying these challenges means specifying aspects of the test construct and hence addresses aspects of theory-based validity.

## 2.2 Error Analysis

Error analysis (EA), as a branch of applied linguistics, is concerned with the study and analysis of errors made by L2 learners. Building on the work of Corder (1974), Ellis and Barkhuizen (2005, 57) distinguish the following five steps in conducting an EA: (1) collecting a sample of learner language, (2) identification of errors, (3) description of errors, (4) explanation of errors, and (5) error evaluation. Steps two and three are of interest in this study and will therefore be explained further.

The second step, *identification of errors*, involves the recognition of elements in the learner's production that deviate from the norm of the L2 in some way. According to Ellis and Barkhuizen (2005), identification "involves a comparison between what the learner has produced and what a native speaker counterpart would produce in the same context" (58). This description builds on the work of Lennon (1991), who defines error as "a linguistic form or combination of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speakers' native speaker counterparts" (182).

An important process here is to "prepare a reconstruction of the sample as this would have been produced by the learner's native speaker counterpart" (Ellis and Barkhuizen 2005, 58). The reconstructed version is referred to as *authoritative reconstruction* of the learner performance. Arriving at such a version is often not without problems. In many cases, the intended meaning of the sentence/utterance is not clear and thus, several reconstructions are possible. For such cases, Corder (1974) suggests seeking an *authoritative interpretation* by asking the learner what they intended to communicate (127–128). This procedure, however, is unpractical for most EA. Therefore, he proposes plausible interpretation as an alternative. Here, the researcher has to "attempt to infer the meaning intended by the learner from the surface structure and his text-sentence in conjunction with the information derived from its context" (128).

The third step, *description of errors*, involves "specifying how the forms produced by the learner differ from those produced by the learner's native speaker counterparts" (Ellis and Barkhuizen 2005, 60). Consequently, the most important process in this step is choosing or developing an error classification system with descriptive categories for coding the errors which have been identified. A system of error categories is referred to as *error taxonomy* (see, e.g., James [1998, ch. 4] for an overview of different approaches).

EA does have some methodological shortcomings and limitations. First, although learners make errors in both comprehension and production, EA can only deal effectively with errors in speaking and writing because errors only manifest themselves visibly in the productive skills. Second, EA gives an incomplete picture of the learning process because it focuses on what the learners do wrongly and not on what they already know



(Hammarberg 1974, 185; Ellis and Barkhuizen 2005, 70; Saville-Troike 2006, 40). Additionally, EA cannot account for learner use of communicative strategies such as avoidance of difficult structures. The absence of errors in a certain area of the language, thus, does not mean that the learner has mastered it (Khansir 2012, 1030; Saville-Troike 2006, 40; Schachter 1974; X. Yu 2017). Third, Lennon (2008) points out problems in assigning the psycholinguistic cause of errors because error explanation is in its nature speculative, and it is often not possible to unambiguously locate the source of an error. Last, there is ambiguity in error classification and identification (Ellis and Barkhuizen 2005, 59–60; Lennon 2008, 55; James 1998, 91–92). The *Scope – Substance* error taxonomy (Dobrić and Sigott 2014) was created as an attempt to alleviate this problem.

Although the research focus of EA has changed, it is still useful as a “methodology for dealing with data, rather than a theory of acquisition” (Cook 1993, 22). This is precisely how EA has been used in studying the impact of learner errors on human ratings of learner performances (e.g., Pibal 2012; Pibal, Sigott, and Cesnik 2018), and this is also the purpose for which it is used in the present study.

### 2.3 The *Scope – Substance* Error Taxonomy

The *Scope – Substance* error taxonomy is an error classification system that was proposed by Dobrić and Sigott (2014) to alleviate problems with subjectivity in existing taxonomies. In this model, errors are classified on the basis of two concepts: *substance* and *scope*. *Substance* is defined as the “the size of the element that needs to be changed in order to correct the error”, while *scope* refers to “the amount of textual or extratextual context that is required for recognising the presence of an error” (114). The idea of characterising errors in this manner was originally proposed by Lennon (1991), as the authors of the taxonomy point out (Sigott, Cesnik, and Dobrić 2016, 80). Similar to the concepts of *scope* and *substance*, Lennon suggested classifying errors by means of their *extent* and *domain*.

The core idea of the *Scope – Substance* error taxonomy is that most errors only become identifiable when there is an incompatibility between a unit in the text (the potential error) and the surrounding context. As Lennon (1991) points out, “most ‘erroneous forms’ are, in fact, in themselves not erroneous at all, but become erroneous only in the context of the larger linguistic unit in which they occur” (189). For example, looking at the word ‘neccessary’ in isolation is enough to see that it is erroneous. However, in the clause ‘I was very nervously’, every word and phrase is in itself acceptable. Only when we consider the context of the whole clause does it become apparent that the use of the adverb ‘nervously’ is not correct.

The size of the constituent in which the error is located and the surrounding context that needs to be checked for incompatibility is expressed in terms of syntactic units. These

are described in terms of units of the commonly accepted grammatical hierarchy, starting with morpheme, the smallest unit, which makes up words, which make up phrases, which make up clauses, which make up sentences. These units are placed in a hierarchy on the basis of their potential size or extensibility and not their actual size or length (referring to the number of constituents) as a sentence may consist of only one clause, or a phrase of only one word (e.g., all three phrases in ‘I am nervous’). The phenomenon that one unit may be the only constituent into which another unit can be analyzed is referred to as unitary constituency (Quirk et al. 1985). In the most extreme case, for instance ‘Bye,’ a structure may be described as a sentence, a clause, a phrase, a word, or a morpheme. The *Scope – Substance* error taxonomy uses the principles of syntactic analysis laid out in Quirk et al. (1985) since a theory-neutral, descriptive grammar is very well suited as a framework for error description (cf. James 1998, 96). In contrast to other, more recent, grammars of English, it is likely that almost every English sentence structure, even learner language, is describable following its principles.

In the taxonomy, errors are described by a combination of *scope* and *substance*. Using the units from the grammatical hierarchy without morpheme but with punctuation and text added, the latest version of the taxonomy has 19 different error types (Sigott, Cesnik, and Dobrić 2016) (see Table 1).

Code	<i>Substance</i>	<i>Scope</i>	Code	<i>Substance</i>	<i>Scope</i>
11	WORD	WORD	34	CLAUSE	SENTENCE
12	WORD	PHRASE	35	CLAUSE	TEXT
13	WORD	CLAUSE	44	SENTENCE	SENTENCE
14	WORD	SENTENCE	45	SENTENCE	TEXT
15	WORD	TEXT	55	TEXT	TEXT
22	PHRASE	PHRASE	92	PUNCTUATION	PHRASE
23	PHRASE	CLAUSE	93	PUNCTUATION	CLAUSE
24	PHRASE	SENTENCE	94	PUNCTUATION	SENTENCE
25	PHRASE	TEXT	95	PUNCTUATION	TEXT
33	CLAUSE	CLAUSE			

Table 1: The 19 error types of the updated taxonomy (with error codes).

Mathematically, 30 combinations would be possible. However, the *scope* of an error cannot be lower than its *substance*, which eliminates 10 combinations like *substance* PHRASE – *scope* WORD. Additionally, *substance* PUNCTUATION – *scope* WORD is not a possible error type (Sigott, Cesnik, and Dobrić 2016).

In Figure 1, the different error types are represented diagrammatically using a coordinate system, showing graphically that the *scope* of an error cannot be lower than its *substance*.

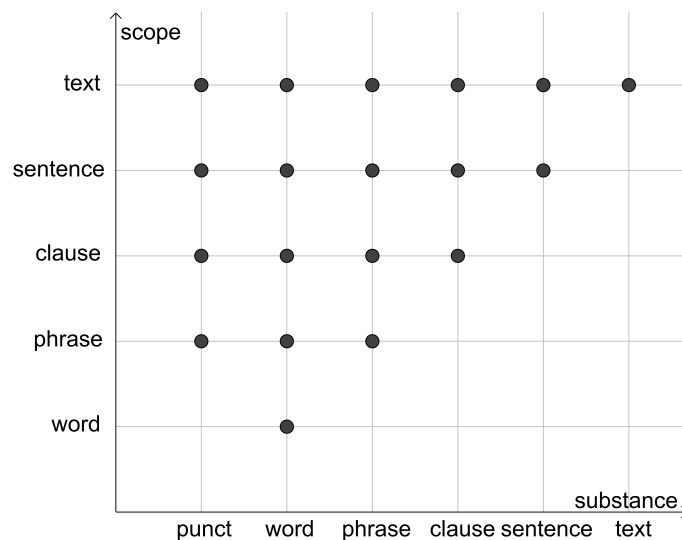


Figure 1: Diagrammatic representation of the 19 error types.

Following are examples that illustrate the application of the taxonomy (for additional examples see Dobrić and Sigott 2014; Sigott, Cesnik, and Dobrić 2016; Hafner 2018a).

#### *substance* word – *scope* word

- (1) If I could fly, I would go to Madrid and get an autogramm from Ronaldo.

In the first example, the word ‘autogramm’ is not part of the English lexicon. In order to repair the error, a correction on the word level must be performed, making the *substance* WORD. The *scope* of the error is WORD since it is sufficient to look at the word in isolation to identify the error.

#### *substance* word – *scope* text

- (2) We go around with them, talk with them, and have a lot of fun.

This sentence does not seem to have any erroneous forms at all when looked at in isolation. However, if the context is widened to the text level, it becomes clear that the learner is talking about events in the past. Therefore, the verbs must be put into past tense. These errors have *substance* WORD and *scope* TEXT.

*substance* phrase – *scope* clause

- (3) When you will phone me, I'll come.

The tense of the verb phrase 'will phone' is not acceptable in a clause that is introduced with 'when'. Hence, the auxiliary verb 'will' needs to be deleted, which changes the structure of the verb phrase (from having an auxiliary slot to not having one). The error becomes visible when the *scope* is widened to the clause level ('When you will phone me'). Consequently, the *substance* is PHRASE while the *scope* is CLAUSE.

*substance* TEXT – *scope* TEXT

This error type typically occurs when the conventions of text structure are violated, such as the logical order of sentences (topic sentence – supporting details – concluding sentence; cause-effect relations, etc.).

- (4) Ten people are on vacation on a boat. The star is Ben Klin Hof. A shark kills one of the people while he was swimming in the sea.

This extract is from a paragraph describing the plot of a movie. In isolation, each of the sentences is acceptable. However, the second one is not appropriate in this slot. Correcting the error involves changing the position of this sentence into a more appropriate one, which changes the text structure. Therefore, both *substance* and *scope* are TEXT.

*substance* PUNCTUATION – *scope* SENTENCE

For the sake of reproducibility, only obligatory commas are considered errors (e.g., lists, initial adverbials that are clauses, non-defining relative clauses, appositions, question tags, introductory comma in informal letters).

- (5) When we arrived there I was very sad.

Since the temporal clause 'When we arrived there' comes before the main clause, the comma between the two clauses is obligatory. The correction thus involves inserting a comma, making the *substance* PUNCTUATION. The error becomes apparent when the whole sentence is considered, thus the *scope* is SENTENCE.

## 2.4 Austrian Educational Standards-Based Writing Test for English at Grade 8 (E8 Standards Writing Test)

The E8 Standards Writing Test is part of a large-scale secondary-level monitoring program of EFL competences at grade 8 in Austria. The main focus of the program is to monitor the listening, reading, writing, and with a restricted sample, speaking abilities

in EFL of pupils in grade 8 and to provide diagnostic information to help transform the teaching and learning of EFL in the Austrian school system (Kulmhofer and Siller 2018). The test takers are approximately 14 years old when they take the E8 Standards Writing Test and should have reached the CEFR levels A2 to B1, depending on the area of competence and descriptor. In 2009, a baseline study was conducted, the first nation-wide Standards Test was administered in 2013, and the most recent one took place in 2019.

In the E8 Standards Writing Test, candidates are assessed in four areas: *Task Achievement*, *Coherence & Cohesion*, *Grammar*, and *Vocabulary*. These areas represent the four dimensions of the analytic rating scale used for the assessment of the writing performances. Each dimension has seven levels plus a '0' band, which is reserved for performances with no assessable language. Four levels (1, 3, 5, 7) have descriptors, while the three levels in between (2, 4, 6) are empty (Gassner, Mewald, and Sigott 2008). The rating scale used for the 2009 baseline study (version May 2008) is reprinted in the appendix.

In the 2009 baseline study, each student received a booklet containing a short and a long answer prompt. The candidates had 10 minutes to produce the short performance and 25 minutes for the long one; time management, however, was left to the students. The prompt instructed the students on the required length of the text and the content points to be addressed. The short answer prompt demanded between 40–60 words and contained between three and five bullet points, and the long answer prompt asked for 120–150 words with five to eight bullet points (Kulmhofer and Siller 2018). Additionally, the writing prompts specified language functions like inviting, apologizing, asking for something or giving advice that the test takers need to perform (Gassner et al. 2011).

### 3 Research Design

#### 3.1 Sample and Population

The sample of written performances came from the 2009 baseline study for the E8 Standards Writing Test, in which a stratified random sample of 10,749 eighth-grade pupils from 204 schools from all over Austria participated (IQS n.d.). The performances were taken from an already existing corpus compiled by Pibal (2012), who selected a random sample of 100 long performances from the overall pool. Pibal only selected long performances (120–150 words in length) because they offer more potential for analysis than the short ones. The texts are based on two different prompts. One instructed the students to write a letter to a friend or relative in which they talk about a recent school trip they undertook. The other required pupils to write a text for a youth magazine and describe their favourite film, music, or book.

Pibal digitized the texts for the purpose of further analysis by transcribing them into text files. From this corpus, a random set of 50 performances was drawn for this research. Each text was assigned a unique number from 1 to 50. Two texts had to be excluded (no. 43 and 45) from the study as they proved to be unsuitable for analysis. The performances were incoherent, full of errors, and the writers' intended meaning was unclear.

### 3.2 Variables

#### 3.2.1 Adjusted Human Ratings (Fair Measures)

After the administration of the E8 Standards Writing Test in 2009, all performances were rated by practising EFL teachers from Austrian schools, who went through a specific rater training program, using a seven-band analytic rating scale (see appendix). In order to be able to minimize the effect of differences in rater behaviour, a rating plan with rater overlap for multiple rating of anchor performances was followed. The raw ratings were then subjected to multifaceted Rasch analysis (MFRA) to adjust for differences in rater severity and task difficulty. The output of this analysis is called *Fair Measure*. A person score based on the aggregated ratings of two performances (long and short) was reported for each participant. Pibal (2012) re-conducted MFRA of the original data on a single-performance basis and calculated five Fair Measure variables: Total Fair Measure, Grammar Fair Measure, Vocabulary Fair Measure, Coherence & Cohesion Fair Measure, and Task Achievement Fair Measure. As the raters did not award an overall rating, the Total Fair Measure is an artefact, aggregated from the Fair Measures on the four rating dimensions. Pibal's single performance-based Fair Measures were used in the present study. Table 2 shows how the variables are named in the analysis. The Fair Measures can take a value between 0 and 7, with 7 being the best rating.

Variable	Variable name
Total Fair Measure	TOT.Fair
Grammar Fair Measure	GR.Fair
Vocabulary Fair Measure	VOC.Fair
Coherence & Cohesion Fair Measure	COH.Fair
Task Achievement Fair Measure	TA.Fair

Table 2: Adjusted human rating (Fair Measure) variables

### 3.2.2 Error Density (ED)

To make the occurrence of errors comparable across performances and relating the errors to the human ratings, error density variables were calculated. The error density is defined as the number of errors per one hundred words. Table 3 gives an overview of the variables, their naming pattern, and their computation.

This was done for two reasons. First, the longer the text, the more potential for error it offers. Second, raters usually take the whole performance as a basis for their judgment. Thus, the same number of errors in a short text will, most likely, be judged more severely than in a longer one.

The error density variables take this into account. For example, the total error density (ED\_TOT) for a 100-word text with 10 errors is 10; for a 200-word text with the same number of errors, the value is 5. Consequently, the error density variables allow a more informative comparison of the occurrence of errors independently of the text length, which would not be possible with the absolute frequencies.

Level	Error density variables	Computation
total	total error density (ED_TOT) (1 variable)	$\frac{\text{total number of errors}}{\text{number of words}} \cdot 100$
<i>substance</i>	<i>substance</i> X density (ED_SuX) (6 variables)	$\frac{\text{number of errors with } \textit{substance} \text{ X}}{\text{number of words}} \cdot 100$
<i>scope</i>	<i>scope</i> Y density (ED_ScY) (5 variables)	$\frac{\text{number of errors with } \textit{scope} \text{ Y}}{\text{number of words}} \cdot 100$

ED = error density.  $X \in \{1, 2, 3, 4, 5, 9\}$  and  $Y \in \{1, 2, 3, 4, 5\}$ .

1  $\triangleq$  WORD, 2  $\triangleq$  PHRASE, 3  $\triangleq$  CLAUSE, 4  $\triangleq$  SENTENCE, 5  $\triangleq$  TEXT, 9  $\triangleq$  PUNCTUATION.

Table 3: Overview of the error density (ED) variables, their naming pattern, and their computation.

### 3.3 Guidelines and Agreement Study

The quality of the linguistic information in any annotated learner corpus depends on the reliability of the error tagging (Díaz-Negrillo and Fernández-Domínguez 2006, 88; Dobrić 2015, 36; Plaban, Pabitra, and Anupam 2008, 58). More generally, the validity of a research study is strongly dependent on the reliability of the data (Brants 2000; Plaban, Pabitra, and Anupam 2008, 58). Granger (2003, 467) describes consistency as one of the key requirements for an annotation system to be fully effective. To achieve consistency, she recommends the elaboration of an error manual with detailed descriptions of the error categories and tagging principles (466–467). A similar point was made by Potter and

Levine-Donnerstein (1999), who emphasize the need for a coding scheme, that is, a set of rules that tell annotators how to code (266–267). Using such a manual or coding scheme should help to ensure that the error coding is the result of a systematic examination with minimal uncertainty and subjectivity which should, as a result, lead to higher levels of reliability. Taking this into consideration, an elaborate set of rules and principles (collectively referred to as guidelines) was developed. These guidelines are the product of a collaborative process involving a group of researchers working with the *Scope – Substance* error taxonomy (Bıdık 2016; Hafner 2018b; Sigott, Cesnik, and Dobrić 2016; Steinkellner 2018) and including an agreement study. For more details on this process and the full guidelines, see Hafner (2018a).

### 3.4 Data analysis

The study utilizes a mixed methods approach. The qualitative aspect is the identification and description of errors in the performances. Correlation and regression analyses represent the quantitative part.

#### 3.4.1 Identification and Description of Errors

Building on the theoretical foundation laid out by Ellis and Barkhuizen (2005), the following three-step procedure was applied to each written performance.

- First, the whole text was read to develop an overall understanding.
- Then, the process of error identification started. This process was stepwise and bottom-up, starting with the smallest unit in the grammatical hierarchy. This means that for every sentence in a performance, the individual words were first screened for errors. The process was then repeated for the phrase, clause, sentence, and text level. The output of this procedure was an authoritative reconstruction of the learner language. The guiding principle was to reconstruct the intended meaning of the learner (plausible interpretation) while changing as little as possible (minimal correction).
- Last, the process of error description started. The errors were coded with ATLAS.ti (version 8). After importing the digitized texts into the software, the error type codes from Table 1 were used to create 19 code labels (e.g., 11, 12, 13, ...) each indicating a combination of *substance* and *scope*. These formed the basis for error tagging. Each data segment constituting an error *substance* was marked with one of these codes. The coding was performed by one of the authors [SH], after consultation with other researchers [GS and FS] also working with the taxonomy, in cases of uncertainty. The annotator analysed each text at least twice. To



ensure reliability and consistency of the application of the *Scope – Substance* error taxonomy, an agreement study was carried out before the annotation process (see section 3.3).

### 3.4.2 Statistical Analyses

After the categorization process in ATLAS.ti 8, the data file was merged with the ratings provided by Pibal (Pibal 2012). The statistical analysis was performed with the statistical software R (version 4.1.0).

Descriptive statistics were used to describe the basic features of the data in the study and to provide simple summaries about the sample. The main aim here was to give an insight into the areas that cause the most difficulty for Austrian learners of English at grade 8 (RQ1).

RQ2 was answered using correlation and regression analysis. To assess the strength of the (statistical) association, Spearman's rank-order correlation coefficient (Spearman's  $\rho$ ) was calculated as not all variables are normally distributed and outliers are present. Pearson's product-moment correlation coefficient (Pearson's  $r$ ) is reported too for comparison as  $\rho$  benchmarks *monotonic* relationships, while  $r$  assesses *linear* ones. In linear relationships, variables tend to move together at a constant, i.e., linear, rate, while in a monotonic relationship, variables tend to move in the same relative direction, but not necessarily at a constant rate.

In short, correlation quantifies the degree to which two variables are related. In contrast, linear regression analysis provides information about the change in the dependent variable when the values in the independent variable(s) change. A simple linear regression model is a mathematical equation that allows us to predict the value of  $Y$  (dependent variable) for a given value  $X$  of the independent variable:

$$Y = b + a \cdot X$$

Thus, regression is used for the mathematical modelling of the relationships between the Fair Measures (dependent variables) and the error densities (independent variables). For this study, we conducted *descriptive* modelling to analyse the impact of the independent variables on the dependent ones without assuming or relying on an underlying causal theory.<sup>1</sup> Since the error densities have skewed distributions and outliers are present in the data, the regression coefficients are estimated using an MM-Estimator, a highly robust and highly efficient estimator (R. R. Wilcox 2012, 499), as implemented in the function `lmrob` in the R-package `robustbase` (version 0.93-7) (Maechler et al. 2021) with

---

<sup>1</sup> See (Shmueli 2010) for an excellent overview of the different kinds of statistical modelling.

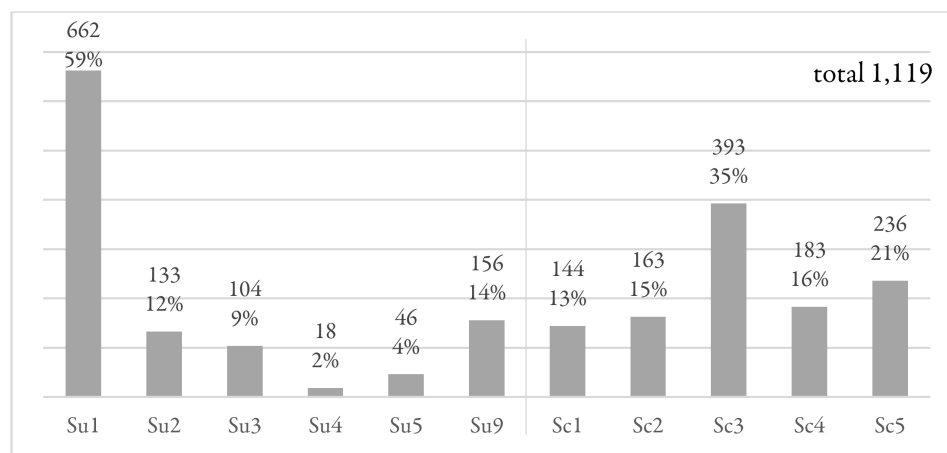
the option `setting="KS2014"` (cf. Koller and Stahel (2011, 2017) and the reference manual of the package on the Comprehensive R Archive Network (CRAN) website). All models were tested for homoscedasticity of residuals (equal variance), autocorrelation, multi-collinearity, and normality of residuals.

## 4 Results

### 4.1 Descriptive Statistics

In total, 1,119 errors were tagged across all 48 texts. On average, every performance contains about 23 errors with a standard deviation of about 10. The median is 22. Virtually all texts contain between 7 and 41 errors and 50% of the performances have between 16 (lower quartile Q1) and 28 errors (upper quartile Q3).

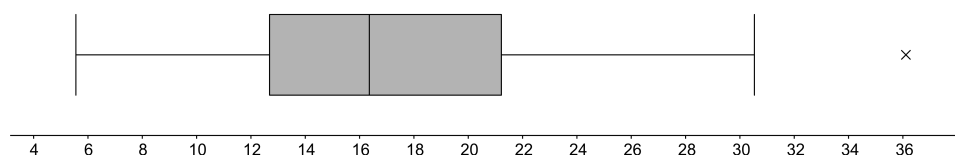
In order to answer RQ1 (*In terms of errors described using the Scope – Substance error taxonomy, what are the most common errors in texts written by Austrian learners of English at grade 8?*), the frequencies of the different *scope* and *substance* errors are shown in Figure 2. *Substance* WORD errors are by far the most frequent (59% of all errors). Larger *substance* errors (CLAUSE, SENTENCE, TEXT) only make up 15% of all errors. Errors with *substance* PUNCTUATION represent the second most common *substance* group (14% of all errors). Errors with a *larger scope* (CLAUSE, SENTENCE, TEXT) constitute almost three quarters (73%) of all errors. *Scope* CLAUSE errors are the most frequent errors among all *scope* categories and constitute 35% of the errors, followed by *scope* TEXT (21%).



SuX = all errors with *substance* X, ScY = all errors with *scope* Y. 1  $\hat{=}$  WORD, 2  $\hat{=}$  PHRASE, 3  $\hat{=}$  CLAUSE, 4  $\hat{=}$  SENTENCE, 5  $\hat{=}$  TEXT, 9  $\hat{=}$  PUNCTUATION

Figure 2: Error frequencies (*scope/substance* categories).

On average, the performances contain 17.03 errors per 100 words with a standard deviation of 6.96. The box plot in Figure 3 displays the distribution of the total error density. The diagram shows that the total error density is approximately normally distributed. In 50% of the performances, the value lies between 12.67 (Q1) and 21.41 (Q3). 50% of the texts have a value of more than 16.36 (median). The minimum is 5.56 and the maximum 36.11. However, the total error density of 36.11 is an extreme value and virtually all texts have between 5.56 and 30.53 errors per 100 words.

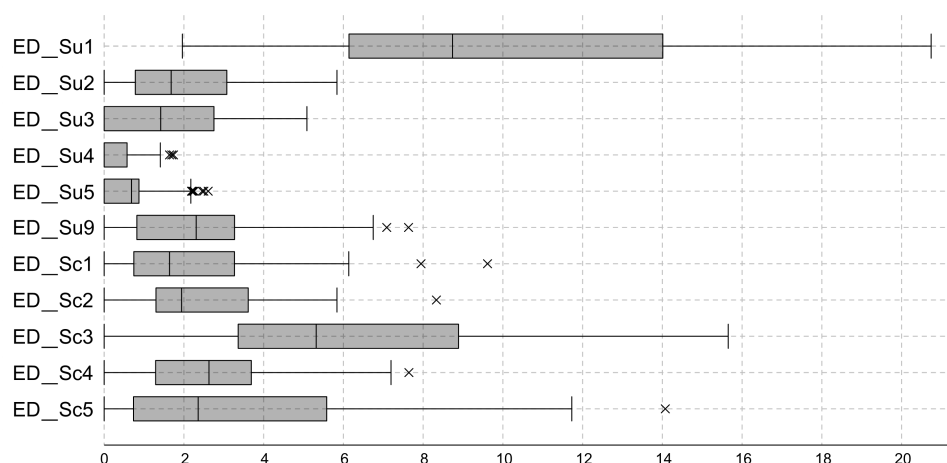


$N = 48$ , min = 5.556, max = 36.111, Q1 = 12.672, median = 16.359, Q3 = 21.412.

The scale refers to number of errors per 100 words. The whiskers have a maximum length of 1.5 times the interquartile range.

Figure 3: Box plot of total error density (ED\_TOT).

The box plots in Figure 4 give more insight into the distribution of the *scope* and *substance* error densities. None of these variables is normally distributed. The tails of the



The scale refers to number of errors (of a specific type) per 100 words. The whiskers have a maximum length of 1.5 times the interquartile range. The extreme value for 'ED\_Su1' with a *substance* word error density of 27.083 is not depicted for reasons of space and layout.

Figure 4: Box plot of all *substance/scope* error densities.

distributions on the right-hand side are longer than on the left-hand side, meaning that the distributions are right-skewed. Most error densities stretch over a wide range of values. Every performance contains errors with *substance* WORD.

## 4.2 Correlation

To answer RQ2 (*What is the relationship between errors described using the Scope – Substance error taxonomy and the human ratings in the four dimensions of Grammar, Vocabulary, Coherence & Cohesion, and Task Achievement?*), the error densities were put into relation to the total fair ratings and to the fair ratings on each of the four dimensions. The relationship was analysed by means of correlation first, and in a following step, by means of regression analysis.

	TOT.Fair		GR.Fair		VOC.Fair		COH.Fair		TA.Fair	
	$\rho$	$r$	$\rho$	$r$	$\rho$	$r$	$\rho$	$r$	$\rho$	$r$
ED_TOT	-.667**	-.618**	-.720**	-.684**	-.569**	-.551**	-.675**	-.602**	-.416**	-.378**
ED_Su1	-.571**	-.561**	-.623**	-.610**	-.513**	-.532**	-.522**	-.501**	-.369**	-.370**
ED_Su2	-.116	-.081	-.204	-.187	-.050	.003	-.252 <sup>+</sup>	-.184	.031	.055
ED_Su3	-.572**	-.527**	-.561**	-.535**	-.453**	-.444**	-.549**	-.540**	-.418**	-.349*
ED_Su4	-.123	-.097	-.118	-.052	-.024	.004	-.168	-.141	-.197	-.144
ED_Su5	-.535**	-.569**	-.471**	-.530**	-.455**	-.498**	-.438**	-.443**	-.554**	-.552**
ED_Su9	.134	.096	.053	.044	.091	.070	.084	.014	.238	.201
ED_Sc1	-.020	-.104	-.076	-.180	-.147	-.190	-.013	-.079	.089	.073
ED_Sc2	-.284 <sup>+</sup>	-.317*	-.280 <sup>+</sup>	-.314*	-.314*	-.325*	-.364*	-.390**	-.157	-.105
ED_Sc3	-.582**	-.528**	-.604**	-.583**	-.517**	-.501**	-.546**	-.506**	-.364*	-.293*
ED_Sc4	-.003	.015	-.105	-.082	-.060	-.008	-.111	-.025	.135	.112
ED_Sc5	-.520**	-.445**	-.465**	-.423**	-.304*	-.275 <sup>+</sup>	-.496**	-.395**	-.532**	-.481**

$N = 48$ . Codes for significance levels (2-tailed): <sup>+</sup>  $p < .1$ . \*  $p < .05$ . \*\*  $p < .01$ .

$\rho$  = Spearman's rank-order correlation coefficient,  $r$  = Pearson's product-moment correlation coefficient

Table 4: Correlation between error densities and Fair Measures.

Table 4 shows the results of the correlation analysis. The following descriptions refer to Spearman's  $\rho$ .

### Total error density

The total error density (ED\_TOT) has highly significant correlations with all Fair Measure variables. The relationships are inverse. Performances with a higher overall fair rating tend to have a lower total error density.

***Substance error densities***

The error densities for *substance* WORD (ED\_Su1), CLAUSE (ED\_Su3), and TEXT (ED\_Su5) have significant, moderate to strong relationships with all Fair Measure variables. The *substance* PHRASE (ED\_Su2), SENTENCE (ED\_Su4), and PUNCTUATION error densities (ED\_Su9) are not significantly correlated with any of the Fair Measures.

***Scope error densities***

The error densities for *scope* CLAUSE (ED\_Sc3) and *scope* TEXT (ED\_Sc5) have moderate to strong correlations with all Fair Measure variables. The *scope* PHRASE error density (ED\_Sc2) has weak correlations with the dimensions Vocabulary and Cohesion & Coherence, has no (significant) connection with the Task Achievement Fair Measure, and only shows a negative trend ( $p < 0.1$ ) for Total and Grammar Fair Measure. ED\_Sc3 has the predominant role for all Fair Measure variables apart from Task Achievement. ED\_Sc5 has the strongest association with this variable. The *scope* WORD (ED\_Sc1) and SENTENCE error densities (ED\_Sc4) are not significantly correlated with any of the Fair Measures.

**4.3 Regression**

In contrast to correlation, which only expresses the strength of the relationship between two variables, regression provides additional information about the change in the dependent variable when the values in the independent variable(s) change. For this reason, we are using the regression coefficients to express sensitivity of the rating system to individual aspects of errors.

outcome	TOT.Fair		GR.Fair		VOC.Fair		COH.Fair		TA.Fair	
model	1		2		3		4		5	
intercept	6.087**	(.394)	6.338**	(.400)	5.925**	(.448)	5.980**	(.454)	5.828**	(.485)
ED_TOT	-.121**	(.021)	-.146**	(.022)	-.111**	(.024)	-.129**	(.025)	-.090**	(.026)
$R^2$	.423		.502		.317		.379		.199	
adj. $R^2$	.411		.491		.302		.366		.181	
robRSE	1.007		1.036		1.149		1.178		1.280	
model	6		7		8		9		10	
intercept	5.350**	(.353)	5.868**	(.443)	5.213**	(.476)	5.514**	(.460)	5.121**	(.398)
ED_Su1	-.091**	(.024)	-.119**	(.030)	-.091**	(.032)	-.099**	(.031)	-.050 <sup>+</sup>	(.027)
ED_Su2	-.036	(.085)	-.126	(.106)	.074	(.114)	-.097	(.110)	.089	(.095)
ED_Su3	-.314**	(.098)	-.289*	(.121)	-.247 <sup>+</sup>	(.130)	-.342*	(.128)	-.101	(.114)
ED_Su4	-.015	(.241)	.149	(.300)	.210	(.323)	-.175	(.313)	-.173	(.270)
ED_Su5	-.445**	(.156)	-.419*	(.196)	-.450*	(.209)	-.260	(.202)	-.889**	(.175)
ED_Su9	.140 <sup>+</sup>	(.071)	.063	(.087)	.095	(.094)	.123	(.094)	.195*	(.080)
$R^2$	.660		.600		.479		.523		.600	
adj. $R^2$	.610		.541		.403		.453		.541	
robRSE	.796		.983		1.062		1.069		.908	
model	11		12		13		14		15	
intercept	5.808**	(.393)	6.088**	(.427)	5.808**	(.497)	5.835**	(.476)	5.318**	(.545)
ED_Sc1	-.014	(.071)	-.057	(.077)	-.061	(.090)	.005	(.088)	.055	(.099)
ED_Sc2	-.145	(.083)	-.159 <sup>+</sup>	(.089)	-.185 <sup>+</sup>	(.105)	-.250*	(.101)	-.052	(.115)
ED_Sc3	-.162**	(.040)	-.182**	(.043)	-.144**	(.050)	-.151**	(.049)	-.115*	(.055)
ED_Sc4	.052	(.078)	-.003	(.085)	.025	(.098)	.024	(.094)	.104	(.108)
ED_Sc5	-.163**	(.041)	-.178**	(.044)	-.105*	(.051)	-.168**	(.050)	-.176**	(.056)
$R^2$	.527		.554		.360		.467		.310	
adj. $R^2$	.471		.501		.284		.414		.228	
robRSE	.958		1.029		1.181		1.137		1.285	

$N = 48$ . Codes for significance levels (2-tailed): <sup>+</sup>  $p < .1$ . \*  $p < .05$ . \*\*  $p < .01$ .

Models in the second column (1, 6, 11) have the outcome variable TOT.Fair, models in the third column (2, 7, 12) GR.Fair, and so on. robRSE = robust residual standard error. Parameter estimation method: MM-estimator. The coefficients are unstandardized, the standard errors are printed in parentheses.

Table 5: Regression models for error densities predicting Fair Measures.

### Total error density

The simple linear regression models for the total error density (ED\_TOT) predicting the Fair Measures are given in models 1 to 5 in Table 5. The relationships are negative, meaning that a higher error density is associated with lower ratings in all dimensions. The total error density has the strongest impact on the Grammar Fair Measure and the weakest on the Task Achievement Fair Measure. The (estimated) regression function from model 2 can be formulated as follows:

$$\text{Grammar Fair Measure: GR.Fair} = 6.338 - 0.146 \cdot \text{ED\_TOT}$$

Using descriptive language, this equation can be described as follows: Comparing two learners of English at grade 8 that differ by 1 in the total error density, the Grammar Fair Measure is expected to differ by 0.146, and the learner with the lower error density has the higher rating. Theoretically, a performance without any errors would have a rating of 6.338.

All models have limited accuracy. The proportion of the variance in the outcome variable Task Achievement Fair Measure that is predictable by the total error density is even as low as 20% ( $R^2 = 0.199$ ). The highest coefficient of determination can be found in the regression model on Grammar Fair Measure ( $R^2 = 0.502$ )

### Substance error densities

The multiple linear regression models between the *substance* error densities (ED\_SuX) and the Fair Measures are shown in models 6 to 10 in Table 5. The error densities for *substance* WORD (ED\_Su1), *substance* CLAUSE (ED\_Su3), and TEXT (ED\_Su5) are significant regressors in almost all models. ED\_Su5 has the largest coefficients. Its impact on the Task Achievement dimension is especially strong. A difference of 1 in the error density is associated with a difference in 0.889 in the Task Achievement rating. In contrast, the error density for *substance* PUNCTUATION (ED\_Su9) has a significant positive coefficient in the Task Achievement model (model 10). This is the only significant positive coefficient.

Similar to the regression with the total error density, all models have limited accuracy. The highest proportion of the variance in the dependent variables that is explained by the *substance* error densities is 66% (model 6). This time, the model with Task Achievement has a relatively high  $R^2$  (0.6).

### Scope error densities

The regression models for the *scope* error densities (ED\_ScY) predicting the Fair Measures are shown in models 11 to 15 in Table 5. The error densities for *scope* CLAUSE

(ED\_Sc3) and TEXT (ED\_Sc5) have a significant negative impact on all Fair Measure variables. The situation is less clear for the *scope* PHRASE (ED\_Sc2) error density. While the coefficients have a relatively large, negative value, the variable is only significant at the 0.05 level (two-tailed) for the Coherence & Cohesion Fair Measure (model 14). However, there is a similar trend ( $p < 0.1$ ) on the other ratings (except for Task Achievement). For none of the five models are the *scope* WORD and SENTENCE error densities statistically distinguishable from zero.

The accuracy of the models is again not high. The highest coefficient of determination is 0.554 in the regression on Grammar Fair Measure (model 12), and the lowest is 0.31 for Task Achievement Fair Measure (model 15).

## 5 Discussion

### 5.1 Areas of Difficulty for Learners

The average total error density is approximately 17, meaning that, on average, each performance contains around 17 errors per 100 words. This leads to the conclusion that writing, especially producing accurate language, is a great difficulty for Austrian learners at grade 8.

The majority of the errors concern the unit WORD, which is not surprising considering that most of the errors relate to prepositions, spelling, personal pronouns, lexical choice and capitalization (cf. Pibal 2012), most of which are correctable by changing a single word. The high frequency of punctuation errors, most of which concern commas, may have two causes. First, the errors may result from L1 interference (language transfer). For example, in German it is required to surround every subordinate clause with commas, which is not the case for English. Additionally, the incorrect addition or omission of a comma usually does not lead to a change in meaning or to a communication problem and, thus, it is reasonable to conclude that students do not pay a lot of attention to (certain) comma rules and the related errors. For the same reasons, teachers might not consider it necessary to give detailed feedback on such errors or provide clear instructions on the correct usage.

Most errors involve *scope* beyond PHRASE (*scope* CLAUSE or SENTENCE constitute 51% of all non-norm adequate forms) and errors with *scope* TEXT turn out to be the second most common category (21%). This leads to the conclusion that a lot of pupils are not yet able to see their text as a meaningful circle of ideas that are interconnected. Thus, recognising the repercussions an error has on textual context seems to be a major challenge for the learners and should be focused on in teaching.



## 5.2 Error Densities and Fair Measures: Implications for Validity

It turns out that a high error density is typical of performances with a low overall rating, thus indicating that errors and writing competence are connected. This finding is not unexpected and has been observed in numerous studies (e.g., Bıdık 2016; Frey and Heringer 2007; Homburg 1984; Weltig 2004; Wolfe-Quintero, Inagaki, and Kim 1998).

In what follows, only the results of the regression analyses are discussed because they provide more information than correlation. While correlation, which only expresses the strength of the relationship between two variables, regression also indicates the change in the dependent variable brought about by a change of the values in the independent variable(s). For this reason, we use the regression coefficients to express sensitivity of the rating system to individual errors.

A meaningful observation concerns the intercept of the regression models. The intercept is the expected mean value of the dependent variable, here the Fair Measure, when all independent variables, here error densities, are zero. In this context, it can be interpreted as the rating of a performance that does not contain any errors. For none of the models is the intercept 7, which would be the perfect rating. It appears that errors alone do not fully explain the rating, which is also evident by the limited regression accuracy.

The total error density has a relatively strong, significant impact on most of the dependent variables. For example, a difference in the value of the total error density of 8 is connected to a difference in one level of the 7-level Total Fair Measure (see appendix). Given that 50% of the performances have more than 16.36 errors per 100 words, this is a considerable impact.

However, the total error density has only a small impact on the rating in the Task Achievement dimension. Task achievement refers to the ability to produce texts that respond to the tasks in a clear and meaningful way and to elaborate and expand ideas meaningfully (Kulmhofer and Siller 2018). Consequently, the total number of errors identified using the *Scope – Substance* error taxonomy should not have a strong impact on this dimension. Since the regression weight is very small ( $\beta = -0.090$ ,  $p < 0.01$ , regression model 5) and the accuracy very low ( $R^2 = 0.199$ ), raters seem to consider different things when making their judgments, such as whether all content points from the prompt have been mentioned or whether they have been elaborated or not (cf. appendix and the extended scale in Gassner, Mewald, and Sigott (2008, 25)).

Among the error features which exert a strong statistical influence on the ratings are the *substance* WORD (ED\_Su1), *CLAUSE* (ED\_Su3) and *TEXT* (ED\_Su5) error densities. *substance* WORD affects ratings on three of the four rating dimensions, more strongly on Grammar than on Vocabulary. *substance* CLAUSE also affects ratings on Grammar and, more strongly, on Cohesion and Coherence. Not surprisingly, *substance* TEXT has the

strongest statistical effect on Task Achievement but also strongly affects the ratings on Grammar and Vocabulary. The effect of *substance* TEXT on Task Achievement is in line with expectations based on the fact that *substance* TEXT errors are the only error type that directly captures problems with task achievement.

Contrary to the expectation that all error types have a negative effect on the rating, the *substance* PUNCTUATION error density has a positive, statistically significant, regression weight for the Task Achievement dimension ( $\beta = 0.195$ ,  $p < 0.05$ , regression model 10). The incidence of punctuation errors, then, increases with text quality. This is presumably due to the fact that the higher the quality of the text, the more complex the structures, but the more punctuation is needed. Conceivably, this increases the writers' chances of making punctuation errors. In contrast to the other error types, punctuation errors, while frequent, do not affect the ratings negatively. A similar phenomenon was observed by Bidik (2016).

Concerning the amount of context needed to detect an error, i.e., *scope*, the error categories CLAUSE (ED\_Sc3) and TEXT (ED\_Sc5) are the ones with a significant impact on the ratings. *Scope* CLAUSE affects all the ratings, but most strongly the Grammar rating, while *scope* TEXT also affects all the ratings but most strongly the Grammar and Task Achievement ratings.

These results constitute detailed information about aspects of student writing that seem to be important in determining quality assessments of the writing. The *substance* categories which are associated with the ratings provide information on what kind of changes in the performances would lead to higher ratings. The *scope* categories could constitute a basis for formulating feedback to students by pointing out to them how much of their text they need to consider in order to avoid particular *substance* errors. Thus, the results could also be seen as a contribution to formative assessment.

In all, the results make a contribution to identifying the construct of the Austrian E8 Standards Writing test. They address context validity by identifying features that the raters seem to be sensitive to, and thus constitute sources of task difficulty. At the same time, the results also speak to theory-based validity as they identify errors that Austrian adolescent writers need to avoid, i.e., areas of competence that they need to develop, in order to enhance their writing.

However, it has to be borne in mind that the error categories studied do not explain the entire variance in the ratings. This indicates that factors other than errors, namely positive features, do seem to play a role in the assessment as well. Identifying such positive features is a worthwhile and necessary task for future research.

## 6 Conclusion

This study analysed errors by means of the *Scope – Substance* error taxonomy in a sample of 50 written performances from the 2009 baseline study for the E8 Standards Writing Test. The primary aim was to contribute to the validation of this large-scale assessment by studying the relationship between errors (described using the *Scope – Substance* error taxonomy) and human ratings awarded to writing performances (RQ2). An additional aim was to examine the data to determine the most common errors in the writing of 14-year-old Austrian learners of English (RQ1).

The study provides insight into the areas in which pupils have difficulties. In general, writing, especially producing accurate language, poses a great difficulty for Austrian learners of English at grade 8. Most of the problems concern the unit **WORD** (59%), followed by **PUNCTUATION** (14%). Regarding *scope*, learners have problems considering the constraints of wider context (clause, sentence, text) to produce accurate language.

These results constitute baseline information about error occurrence in adolescent L2 English writing of L1 German-speaking learners. Unlike in the US, where data on error occurrence has been monitored over decades (A. A. Lunsford and K. J. Lunsford 2008; K. C. Wilcox, Yagelski, and F. Yu 2013), such information has not been collected for L2 English writing in the German-speaking context. Now that this study has established the baseline, it would be interesting to follow the development of error incidence in L2 English writing in the German-speaking area in the future.

The data indicate that errors and assessments of writing competence in the E8 Standards Writing Test are connected. In line with prior studies, this study also showed that as writers become more proficient, they tend to produce increasingly accurate language. In this study, this manifests itself as a negative relationship between human ratings and the presence of errors identified by means of the *Scope – Substance* error taxonomy. A low error density is associated with high ratings and a high error density with low ratings. *substance* **WORD**, **CLAUSE**, and **TEXT** error densities play an important role in the rating in most dimensions; errors with a larger *scope* also have a strong effect. By highlighting aspects of errors to which raters seem to be sensitive, these findings constitute evidence of context validity. At the same time, the findings are relevant to theory-based validity by concretising areas of competence that learners need to develop in order to receive higher ratings. While errors are important determinants of the ratings, additional factors must be at play as the accuracy of the regression models is not perfect. It seems that raters do take errors into consideration in the assessment, but other variables, presumably positive features, also contribute to their decisions. This should in fact be the case since the rating scale refers to negative as well as positive features.

The study has implications for practice. On the one hand, it provides a basis for refining the descriptors of the E8 rating scale for writing by adding the aspects of error to which raters seem to be particularly sensitive. This could help to make the rating easier and contribute to reliability and validity. Making these aspects of errors explicit in rater training should also prove helpful in efforts to maximise reliability and validity. On the other hand, if aspects of errors to which raters are sensitive are communicated to teachers, the competences required to avoid such errors can be focused on in the teaching of writing. This can facilitate positive washback, which is an important aim of Austrian Educational Standards testing.

## References

- Bıdık, Buket. 2016. "Refining the SD Error Taxonomy: A Descriptive Analysis of Grammatical Errors in Samples by Turkish EFL Learners." MA thesis, Department of English and American Studies, University of Klagenfurt.
- Brants, Thorsten. 2000. "Inter-Annotator Agreement for a German Newspaper Corpus." In *Second International Conference on Language Resources and Evaluation LREC-2000*. Athens, Greece. <http://www.coli.uni-saarland.de/~thorsten/publications/Brants-LREC00.pdf>.
- Cizek, Gregory, and Michael Bunch. 2007. *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*. Thousand Oaks, California: SAGE Publications, Inc.
- Cook, Vivian J. 1993. *Linguistics and Second Language Acquisition*. Basingstoke: Macmillan.
- Corder, Pit. 1974. "Error Analysis." In *Techniques in Applied Linguistics*, edited by John P. B. Allen and Pit Corder, 122–154. The Edinburgh Course in Applied Linguistics 3. London: Oxford University Press.
- Díaz-Negrillo, Ana, and Jesús Fernández-Domínguez. 2006. "Error Tagging Systems for Learner Corpora." *RESLA* 19: 83–102.
- Dobrić, Nikola. 2015. "Quality Measurements of Error Annotation - Ensuring Validity Through Reliability." *The European English Messenger* 24 (1): 36–42.
- Dobrić, Nikola. 2020. "Rater Cognition and Errors: A Corpus-Based Approach to Validating Writing Assessment." Habilitationsschrift, University of Klagenfurt.
- Dobrić, Nikola, and Günther Sigott. 2014. "Towards an Error Taxonomy for Student Writing." *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 19 (2): 111–118. <http://tujournals.ulb.tu-darmstadt.de/index.php/zif/article/download/35/32>.
- Dobrić, Nikola, Günther Sigott, Gašper Ilc, Vesna Lazović, Hermann Cesnik, and Andrej Stopar. 2021. "Errors as Indicators of Writing Task Difficulty at the Slovene General Matura in English." *International Journal of Applied Linguistics*, 1–17. <https://doi.org/10.1111/ijal.12345>.
- Ellis, Rod, and Gary Barkhuizen. 2005. *Analysing Learner Language*. Oxford applied linguistics. Oxford: Oxford University Press.
- Fliri, Benjamin, ed. 2018. *ÖGSD Tagungsberichte Vol. 4*: 10. Nachwuchstagung. Sprachendidaktik: Der Wissenschaftliche Nachwuchs Im Dialog. (Proceedings of the 10th ÖGSD Young

- Researchers' Conference). Graz: ÖGSD. <https://www.oegsd.at/wp-content/uploads/2020/08/2018-Nachwuchstagung-Bericht-und-extended-Abstracts.pdf>.
- Frey, Evelyn, and Hans Jürgen Heringer. 2007. "Automatische Bewertung Schriftlicher Lernerproduktionen." *Linguistische Berichte* 211: 331–345.
- Gassner, Otmar, C. Mewald, R. Brock, F. Lackenbauer, and Klaus Siller. 2011. *Testing Writing for the E8 Standards: Technical Report 2011*. Salzburg: BIFIE Salzburg. Accessed February 08, 2018. [http://www.bifie.at/wp-content/uploads/2017/05/bist\\_Technical-Report2-E8\\_2011-09-26.pdf](http://www.bifie.at/wp-content/uploads/2017/05/bist_Technical-Report2-E8_2011-09-26.pdf).
- Gassner, Otmar, C. Mewald, and Günther Sigott. 2008. "Testing Writing: Specifications for the E8-Standards Writing Tests." LTC Technical Report 4. Accessed March 07, 2020. [https://www.aau.at/wp-content/uploads/2018/03/LTC\\_Technical\\_Report\\_4.pdf](https://www.aau.at/wp-content/uploads/2018/03/LTC_Technical_Report_4.pdf).
- Granger, Syviane. 2003. "Error-Tagged Learner Corpora and CALL: A Promising Synergy." *CALICO Journal* 20 (3): 465–480.
- Hafner, Samuel. 2018a. "Analyzing Errors in Written Performances by Means of the Scope – Substance Error Taxonomy and Investigating Their Influence on Human Ratings." Diploma thesis, Department of English, University of Klagenfurt. <https://netlibrary.aau.at/urn:nbn:at:at-ubk:1-36593>
- Hafner, Samuel. 2018b. "Refining the Scope – Substance Error Taxonomy by Means of an Agreement Study." In *ÖGSD Tagungsberichte Vol. 4: 10. Nachwuchstagung. Sprachendidaktik: Der Wissenschaftliche Nachwuchs Im Dialog*, edited by Benjamin Fliri et al., 18–20. Graz: ÖGSD.
- Hammarberg, Björn. 1974. "The Insufficiency of Error Analysis." *International Review of Applied Linguistics in Language Teaching* 12: 185–192.
- Homburg, Taco Justus. 1984. "Holistic Evaluation of ESL Compositions: Can It Be Validated Objectively?" *TESOL Quarterly* 18 (1): 87–107. <https://doi.org/10.2307/3586337>.
- IQS. n.d. "Ausgangsmessung, 8. Schulstufe (2009)." Accessed May 18, 2021. <https://www.iqs.gv.at/themen/bildungsforschung/forschungsdatenbibliothek/daten-der-bildungsstandardueberpruefungen/fdb-ausgangsmessung-8-schulstufe-2009>.
- James, Carl. 1998. *Errors in Language Learning and Use: Exploring Error Analysis*. Applied Linguistics and Language Study. London, New York: Longman.
- Khansir, Ali Akbar. 2012. "Error Analysis and Second Language Acquisition." *Theory and Practice in Language Studies* 2 (5): 1027–1032. <https://doi.org/10.4304/tpls.2.5.1027-1032>.
- Koller, Manuel, and Werner A. Stahel. 2011. "Sharpening Wald-Type Inference in Robust Regression for Small Samples." *Computational Statistics & Data Analysis* 55 (8): 2504–2515. <https://doi.org/10.1016/j.csda.2011.02.014>.
- Koller, Manuel, and Werner A. Stahel. 2017. "Nonsingular Subsampling for Regression S Estimators with Categorical Predictors." *Computational Statistics* 32 (2): 631–646. <https://doi.org/10.1007/s00180-016-0679-x>.
- Kulmhofer, Andrea, and Klaus Siller. 2018. "The Development of the Austrian Educational Standards Test for English Writing at Grade 8." In *Language Testing in Austria: Taking Stock/Sprachtesten in Österreich: Eine Bestandsaufnahme*, edited by Günther Sigott, 129–145. Frankfurt am Main: Peter Lang.

- Lennon, Paul. 1991. "Error: Some Problems of Definition, Identification, and Distinction." *Applied Linguistics* 12 (2): 180–196. <https://doi.org/10.1093/applin/12.2.180>.
- Lennon, Paul. 2008. "Contrastive Analysis, Error Analysis, Interlanguage." In *Bielefeld Introduction to Applied Linguistics*, edited by S. Gramley and V. Gramley, 51–60. Bielefeld: Aisthesis.
- Lumley, Tom. 2005. *Assessing Second Language Writing: The Rater's Perspective*. Berlin: Peter Lang.
- Lunsford, Andrea A., and Karen J. Lunsford. 2008. "'Mistakes Are a Fact of Life': A National Comparative Study." *College Composition and Communication* 59 (4): 781–806. [https://www.jstor.org/stable/20457033?seq=1#metadata\\_info\\_tab\\_contents](https://www.jstor.org/stable/20457033?seq=1#metadata_info_tab_contents).
- Maechler, Martin, Peter Rousseeuw, Croux, Christophe, Todorov, Valentin, Andreas Ruckstuhl, Matias Salibian-Barrera, Tobias Verbeke, Manuel Koller, Eduardo L. T. Conceicao, and Maria Anna Di Palma. 2021. *Robustbase: Basic Robust Statistics R Package Version 0.93-7*. <https://CRAN.R-project.org/package=robustbase>.
- Pibal, Florian. 2012. "Identifying Errors in the Written Manifestations of Austrian English Learner Language at 8<sup>th</sup>-Grade Secondary Level and Their Influence on Human Ratings." MA thesis, Department of English and American Studies, University of Klagenfurt.
- Pibal, Florian, Günther Sigott, and Hermann Cesnik. 2018. "The Role of Error in Assessing Writing in the National Educational Standards Baseline Test." In *Language Testing in Austria: Taking Stock/Sprachtesten in Österreich: Eine Bestandsaufnahme*, edited by Günther Sigott, 419–443. Frankfurt am Main: Peter Lang.
- Plaban, Bhowmick, Mitra Pabitra, and Basu Anupam. 2008. "An Agreement Measure for Determining Inter-Annotator Reliability of Human Judgements on Affective Text." In *Proceedings of the Workshop on Human Judgements in Computational Linguistics*, edited by Ron Artstein, Gemma Boleda, Frank Keller, and Sabine Schulte im Walde, 58–65. Manchester, UK: Coling 2008 Organizing Committee. <http://aclweb.org/anthology/W08-12>.
- Potter, W. James, and Deborah Levine-Donnerstein. 1999. "Rethinking Validity and Reliability in Content Analysis." *Journal of Applied Communication Research* 27 (3): 258–284. <https://doi.org/10.1080/00909889909365539>.
- Quirk, Randolph, Sidney Greenbaun, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Saville-Troike, Muriel. 2006. *Introducing Second Language Acquisition*. Cambridge introductions to language and linguistics. Cambridge: Cambridge University Press.
- Schachter, Jacquelyn. 1974. "An Error in Error Analysis." *Language Learning* 24 (2): 205–214. <https://doi.org/10.1111/j.1467-1770.1974.tb00502.x>.
- Shmueli, Galit. 2010. "To Explain or to Predict?" *Statistical Science* 25 (3): 289–310. <https://doi.org/10.1214/10-STS330>.
- Sigott, Günther. 2004. *Towards Identifying the C-Test Construct*. Frankfurt am Main: Peter Lang.
- Sigott, Günther, ed. 2018. *Language Testing in Austria: Taking Stock/Sprachtesten in Österreich: Eine Bestandsaufnahme*. Frankfurt am Main: Peter Lang.
- Sigott, Günther, Hermann Cesnik, and Nikola Dobrić. 2016. "Refining the Scope – Substance Error Taxonomy: A Closer Look at Substance." In *Corpora in Applied Linguistics: Current Ap-*

- proaches*, edited by Nikola Dobrić, Eva-Maria Graf, and Alexander Onysko, 79–94. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.
- Steinkellner, Florian. 2018. “Towards a Refined Version of the Scope – Substance Error Taxonomy.” In *ÖGSD Tagungsberichte Vol. 4: 10. Nachwuchstagung. Sprachendidaktik: Der Wissenschaftliche Nachwuchs Im Dialog*, edited by Benjamin Fliri et al., 46–48. Graz: ÖGSD.
- Weir, Cyril J. 2005. *Language Testing and Validation: An Evidence-Based Approach*. Research and practice in applied linguistics. Basingstoke: Palgrave Macmillan.
- Weltig, Matthew S. 2004. “Effects of Language Errors and Importance Attributed to Language on Language and Rhetorical-Level Essay Scoring.” *Spain Fellow Working Papers in Second or Foreign Language Assessment* 2: 53–81.
- Wilcox, Kristen Campbell, Robert Yagelski, and Fang Yu. 2013. “The Nature of Error in Adolescent Student Writing.” *Read Writ* 27 (6): 1073–1094. <https://doi.org/10.1007/s11145-013-9492-x>.
- Wilcox, Rand R. 2012. *Introduction to Robust Estimation and Hypothesis Testing*. 3rd ed. Statistical Modeling and Decision Science. Amsterdam: Elsevier.
- Wolfe-Quintero, Kate, Shunji Inagaki, and Hae-Young Kim. 1998. *Second Language Development in Writing: Measures of Fluency, Accuracy, & Complexity*. Technical report / Second Language Teaching & Curriculum Center 17. Honolulu, Hawaii: University of Hawai’i Press.
- Yu, Xiu. 2017. “On the Avoidance Phenomenon in Writing.” *Journal of Language Teaching and Research* 8 (5): 948–952. <https://doi.org/10.17507/jltr.0805.15>.

## Appendix

### E8 Writing Rating Scale (version May 2008)

	Task Achievement	Coherence and Cohesion	Grammar	Vocabulary
7	<ul style="list-style-type: none"> <li>complete task achievement</li> <li>meets text type requirements</li> </ul>	<ul style="list-style-type: none"> <li>cohesive on both sentence and paragraph level</li> <li>clear, coherent text</li> </ul>	<ul style="list-style-type: none"> <li>good range of structures</li> <li>few inaccuracies</li> </ul>	<ul style="list-style-type: none"> <li>good range of vocabulary</li> <li>generally accurate with some incorrect words</li> </ul>
6				
5	<ul style="list-style-type: none"> <li>good task achievement</li> <li>few inconsistencies in text type requirements</li> </ul>	<ul style="list-style-type: none"> <li>good sentence level cohesion</li> <li>some paragraph level coherence and cohesion</li> </ul>	<ul style="list-style-type: none"> <li>generally sufficient range of structures for familiar contexts</li> <li>occasional inaccuracies</li> <li>message clear</li> </ul>	<ul style="list-style-type: none"> <li>sufficient range of vocabulary, communicating clear ideas</li> <li>occasionally inaccurate</li> </ul>
4				
3	<ul style="list-style-type: none"> <li>sufficient task achievement</li> <li>some inconsistencies in text type requirements</li> </ul>	<ul style="list-style-type: none"> <li>some simple sentence level cohesion</li> <li>frequent lack of paragraph level coherence and cohesion</li> </ul>	<ul style="list-style-type: none"> <li>limited range of simple structures</li> <li>frequently inaccurate</li> <li>generally without causing breakdown</li> </ul>	<ul style="list-style-type: none"> <li>limited range of vocabulary, mostly communicating clear ideas</li> <li>frequently inaccurate vocabulary</li> <li>tendency to lift phrases from prompt</li> </ul>
2	<ul style="list-style-type: none"> <li>some task achievement</li> <li>does not meet text type requirements</li> </ul>	<ul style="list-style-type: none"> <li>extremely limited cohesion on sentence and paragraph level</li> <li>text not coherent</li> </ul>	<ul style="list-style-type: none"> <li>extremely limited range of structures</li> <li>mostly inaccurate</li> <li>frequent breakdown of communication</li> </ul>	<ul style="list-style-type: none"> <li>extremely limited range of vocabulary, communicating few clear ideas</li> <li>mostly inaccurate vocabulary</li> <li>several chunks lifted from prompt</li> </ul>
1				
0	no task achievement	no assessable language	no assessable language	no assessable language

Figure 5: E8 Writing Rating Scale (version May 2008) (Gassner, Mewald, and Sigott 2008, 24)



# Conceptualisation of Sustainability

## A Theoretical and Methodological Framework for Understanding Metaphors and Narratives of Sustainability

**Franziska WEDER**

University of Queensland (Australia)

**Nikola DOBRIĆ**

University of Klagenfurt (Austria)

### Abstract

Metaphors are language phenomena commonly used as tools of persuasion, as evident in different kinds of public discourses, most notably in political addresses. They are particularly potent in this respect because they function on the principle of connecting the logical with the emotional. This persuasive role has also been attested in media reports, usually employed for the purposes of framing and the creation of a specific narrative. The purpose of this paper is to outline a theoretical and methodological account of metaphor choices related to sustainability and restoration, as issues related to climate change, made by the contemporary media and the

perceptions formed by repetition and reinforcements of certain kinds of imagery present in their choice. An additional purpose is to understand which information sources, and possible instances of influence and leverage, could be of importance in terms of the media reporting on sustainability-related issues. Therefore, the paper offers novel conceptual and analytical guidelines for future research in the field of sustainability communication.

**Keywords:** *Framing, Sustainability, Conceptual metaphors, Climate change, Media, Narratives*

(c) Franziska Weder & Nikola Dobrić; f.weder@uq.edu.au; nikola.dobric@aau.at

Colloquium: New Philologies, Volume 6, Issue 2 (2021)

doi: 10.23963/cnp.2021.6.2.3

Stable URL: <https://colloquium.aau.at/index.php/Colloquium/article/view/159>

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

## 1 Introduction

Sustainability is at its core a narrative enterprise (Herrick and Pratt 2013), stimulating public discourses in many ways. It is directly related to climate change, focusing on a variety of problematic social issues like species extinction, violation of human rights or diminished biodiversity. As such, it is communicated to the general public in two ways. On one side, the media operate at a “peak negativity” (Atanasova and Flottum 2019), with a strong problematisation angle. This is evident, for example, in their use of war-related metaphors such as “carbon tax proposal battles” or “eco-warriors”. On the other side, in corporate and political discourses, sustainability is used to present potential solutions to reach a ‘better’ future.

Linguistic scholars, literary theorists, and philosophers of language analyse the ways in which language constructs and, by extension, enables social representations and practices. The constitutive role of language opens up a pathway to sustainability, which Herrick and Pratt (2013, 4433) term as depending “upon acceptance of a transformative or constitutive narrative”. The assumption is that any kind of a narrative construct is subject to review, critique, deconstruction, and reconstruction or to rigorous (re)formulation. In this paper, the idea is to combine the complementary perspectives of linguistic and communication studies to deconstruct sustainability communication in public communication, concretely ‘the media’, from a content perspective. In essence, we take stories and narratives as the foundation of meaning and sense making processes (van der Leeuw 2019, Weder et al. 2019, Eisenstein 2013), and focus on metaphors at the core of the sustainability narratives, bending via scientific reasoning as one part of the sustainability story (Frank 2017) to consumer centred, emotive communication and empty “buzz-wording” (Krainer and Weder 2011). This is embedded in a narrative approach to sustainability as an ‘emergent quality’ in relation to the ecological crisis (Sahinidou 2016) and the wider story of climate change.

Metaphors are widely used in public communication, influencing the shape and twist of a narrative. One of the major aims of metaphors in general is to connect the logical (logos) with the emotional (pathos) (Mio 1997, 122). For example, if we observe Extinction Rebellion’s ‘fighting’ climate change, and by the same time being described as ‘Eco-Warriors’, then the intention to resonate with the audience on an emotional level is quite clear from the idiomatic nature of the statement. By introducing a concept of WAR (‘Eco-warriors’) into the related public discourse, a certain course of action is advocated for, for example fighting against climate change<sup>1</sup>. Thus, when people repeatedly hear that

---

<sup>1</sup> Interestingly enough, ‘fighting against climate change’ is another case of a metaphor built on the concept of WAR.

we have to fight climate change rather than calmly ponder on solutions, they are likely to incorporate this very specific ideological model in their mind and, presumably, take a more active approach in this respect.

Viewing the use of language in general and metaphors in particular for manipulating public perception and opinion via political discourse, where their role has long been recognised as crucial, as parallel to the way media present different issues, one is drawn away from cognitive linguistics towards the theoretical concepts of *agenda setting* and *framing* research, within the field of media studies. For instance, when writing about climate change, Fox News chooses to run the story under the following headline: “Let’s chill out about global warming” (January 12, 2017). If we focus on the idiomatic ‘chill out’, the attempt by this media outlet to ‘frame’ climate change as a segment of someone’s irrational behaviour becomes rather evident. The imagery evoked by the conceptual mapping (‘LACK OF MOTION IS COOL TEMPERATURE’) and the associated word play (grounded in the semantic contrast between cool and warm) illustrated by the given headline is meant to evoke an ideological model of climate change being false and nothing to be excited about. Agenda setting research tries to account for such behaviour of media outlets by linking it to the process of thematisation related to the effect media coverage can have in defining the relevance of an issue (Lang and Lang 1959, McCombs and Shaw 1972), while it also seeks to identify interdependencies of issues between different systems of relating information. Here, we can turn to the concept of framing as second-level agenda setting (McCombs and Shaw 1972, McCombs et al. 1997, Weaver 2007) as it is often more accurate to speak of *intermedia agenda setting* (defined as the influence between different types of media and sources of information (Roberts and McCombs 1994, McCombs 2005)). For example, if we look at intermedia agenda setting in the field of health and illness communication and the related news coverage, we can see that government and official sources rank amongst the most reliable and most commonly used information subsidies, most likely due to their perceived credibility and authority (Lacy and Coulson 2000, Sweetser and Brown 2008, Dobrić and Weder 2015). In fact, we can say that that government and official sources ‘organise’ (Weder 2012) and hereby influence the media agenda in health communication and, more importantly and by extent, the public perception of the issue at hand in a broader sense.

## 2 Theoretical Background

From a core scientific point of view, anthropogenic climate change is mainly debated and studied in environmental sciences, natural sciences, and economics. Sustainability communication is at the core of the intersections between the named disciplines and has an interdisciplinary as well as transdisciplinary character (Godemann and Michelsen 2011). On an interdisciplinary plane, however, media and communication studies and linguistics (among many other scientific disciplines) offer a valuable additional perspective, both theoretical and methodological. From a conceptual theoretical perspective, the link between sustainability and metaphors can be located at the intersection of intermedia agenda setting, framing, and cognitive linguistics, and gets its theoretical, as well as conceptual, inspiration from the current state of the art in environmental communication studies and linguistics.

The theoretical concept offered here, aimed to aid in understanding and researching sustainability communication, is embedded in a critical perspective that is typical for environmental communication studies, which has the character of an ‘activist discipline’. Environmental communication as a research field goes beyond the information or transfer of knowledge approach; the potential to shape environmental and societal development and change is a constitutional element of environmental communication studies. Therefore, the field is also described as ‘ecological discourse’, “with the sustainability concept being the most recent communicative framework” from a media and communication perspective (Adomßent and Godemann 2013, 28). The related theoretical concept is further elaborated in the following subsections and is additionally linked to a relevant methodological framework, all in the effort to reflect on the communication about, of, and for sustainability that potentially complements current environmental communication research. Thus, from our perspective, sustainability communication is influenced by the critical character of environmental communication research and scholarship. This is further highlighted by its focus on social representations of nature, human-nature relationships, resource-related behaviour, and public interest issues (such as climate change).

### 2.1 (Intermedia) Agenda Setting and Framing of Sustainability

The concept of agenda setting refers to the capability of the mass news media to not only influence how people think about certain topics but, more importantly, what they perceive as an issue in the first place (McCombs 2004). The suggestion is that within a modern society the mass news media constitute the public agenda, comprising of all issues that at least achieved awareness within the majority of the public (Cobb et al. 1976). It can be argued that the mass news media help to structure and impart the most relevant

information to the public (see Tab. 1), which could more critically also be perceived as manipulation (Rössler 1997). For example, early studies by Parlour and Schatzow (1978) have shown that there is a direct correlation between environmental concerns and the relative amount of news media coverage of them. More drastically, they have found no evidence for public concerns about environmental issues which were not covered by the news media. Therefore, it can be argued that it is through the social orientation function of the media that resonance for the relevance of sustainability can be created in a society (de Witt 2011, Ziemann 2011).

Within the field of sustainability communication, climate change (and now increasingly climate crisis or climate emergency) has been the flagship of environmental and societal (public interest) issues among politics, media, and scientists alike since the late 1980s (Moser 2010, Nerlich et al. 2010). Numerous published articles and extensive mass news media coverage enabled climate change to become a salient issue on the public agenda (Newig 2011). However, the studies are limited to analysing only the structuration and relevance of a certain issue. A research paradigm focusing on framing as a possibility to investigate the creation and character of the meaning of related transmitted messages is needed to complement the current approaches.

Therefore, as Table 1 breaks it down, we introduce frames as persistent patterns of cognition, interpretation, presentation, selection, emphasis, and exclusion, by which symbol-handlers (such as the media) routinely organise public issues (Gitlin 1980). Frames as an organising principle of communication (Weder 2012) have to be linked to the process of framing, a process which “symbolically to meaningfully structure[s] the social world” (Reese, 2001, 5). Despite studies focusing on holistic or generic frames (Gerhards and Rucht 1992, Snow and Benford 1992), our paradigm connects with the majority of approaches working with issue-specific frames (de Vreese 2005), whereby it is assumed that each subject has different thematic frames (Shah et al. 2002). An issue-specific frame can be interpreted as a position or argument based on an opinion. In this respect, issue-specific frames are at the heart of narratives, they are the essence of an issue (Gamson and Modigliani 1989, 3).

	Definition	Sustainability communication
(Public) Discourse	conversations about a certain (public) issue	economic discourse about the changing climate and the transition to renewable energy (with various narratives driving the discourse)
Agenda Setting	mass news media help to structure and impart the most relevant information to the public	The bushfires in Australia did not affect people in Europe directly; however, it was on the media agenda and people were talking about it all over the globe.
Narrative	storyline of an issue, structuring the agenda	There were two narratives about the bushfires: 1) The extent of the fires was related to climate change (natural hazard); 2) The extent of the fires was not related to climate change but to unsustainable management of bush, forests and agriculture.
Frame	position or argument based on an opinion, at the heart of narratives, the essence, constitutive and organising element of an issue; organising principles of a narrative	i.e. ecological disaster, protection of resources, economic threat; restoration; fight against climate change
Conceptual Metaphor	communicative moments employed to transfer a specific meaning, to organise and emphasise it, and to create larger sets of related meaning	i.e. war (we have to fight the fires, fight climate change, eco-warriors); sick ecosystem (lungs of the planet)

Table 1: Narratives, frames and metaphors, theoretical background and definitions

At this point, we can introduce *narratives* as the storylines that lay above frames and that express the issues or cover certain topics. At the same time, a narrative is always set in a particular cultural context (Abbott 2008); a narrative represents a cultural framework and offers a specific understanding of the issue at hand. Frames, subsequently, are the *organising principles of a narrative* (Fig. 1). Therefore, in their turn, they help to identify the nature of the narrative from an analytical point of view.

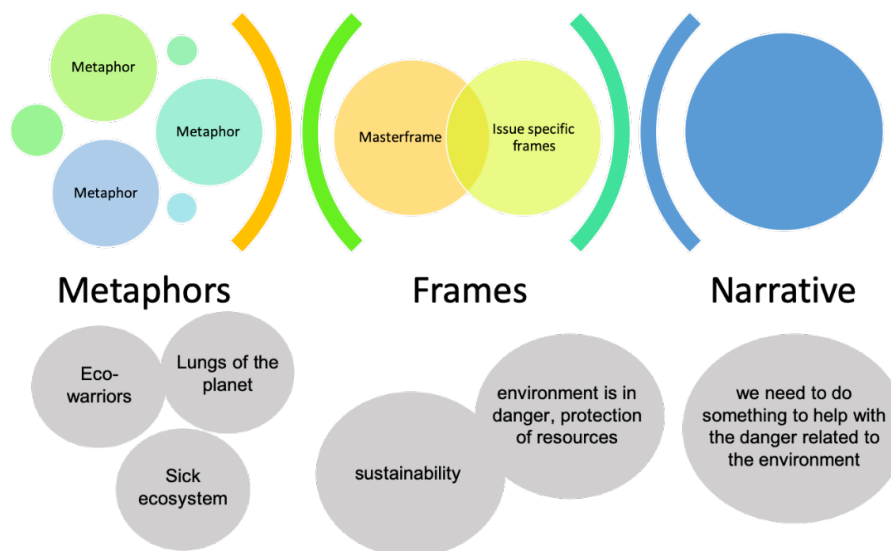


Figure 1: Narratives, frames, and metaphors; framework adapted from Weder (2012).

The narrative of restoration and care about resources needed to meet problems related to climate change and to solve them is represented in public communication. The related frame is *sustainability as normative framework* or, to be more issue-specific, that our environment is in danger and resources have to be protected. There are other issue-specific frames regarding conversations about nature that include protection or the necessity to *fight* against climate change. However, our approach takes this an additional, innovative step further, introducing a yet insufficiently investigated concept of the 'metaphor' as a major 'part' of frames (and, subsequently, public agendas and narratives structuring the agenda). In other words, our focus is on metaphors as *communicative moments* employed to transfer a specific meaning, to organise and emphasise it, and to create larger sets of related meaning (see again Fig. 1).

Here, the potential of metaphors to express and emphasise frames becomes obvious. The concept of WAR introduced in the beginning as grounds of metaphorical expressions, showed us that the terminology of war can be used in communicating the narrative of our natural habitat being threatened and as humans us having to fight against our annihilation. This will be further elaborated in the following section, which complements the already established concepts of media and communication studies through a cognitive linguistic perspective.

## 2.2 Cognitive Linguistics and Conceptual Metaphor Theory

Studies focusing on metaphors in general and on conceptual metaphors in particular, constitute the major strand of cognitive linguistics research (Johnson and Lakoff 1980; Lakoff 1993, Ortony 1993; Steen 1999; Kövecses 2002, 2005; Evans 2004; Gibbs 2006, 2008; Glucksberg and Haught 2006; Glucksberg 2008; to name a few). In short, cognitive linguistics, or more precisely conceptual metaphor theory (CMT), examines metaphors as cognitive phenomena rather than only as lexical ones (or stylistic, as understood in literary terms) (Lakoff and Johnson 1980, 31-50). Metaphors are considered as means of shaping not only the way we speak but the way we think and perceive the world around us, since most of the time we preconsciously use and understand metaphorical language. The process of constructing meaning using metaphorical concepts is called *metaphorisation* and is “founded on association [and it] constructs systems based on prototypical notions and meanings which are used to classify the real world” (Grković-Mejdžor 2008, 54). Metaphors are constructed out of patterns that transcend the individual lexical item, where metaphorical concepts are open-ended sets. The patterns themselves can be grasped as a *mapping* between two domains of experientially grounded knowledge (Evans 2005), the mapping taking the form of a conceptualising of a *target conceptual domain* in terms of the *source conceptual domain* as an alignment between aspects of both source and target (Dobrić 2011). Most commonly, the structure of ‘concrete’ source domains is mapped (Johnson and Lakeoff 1980 252) onto more ‘abstract’ target domains, where the meaning retains the semantic markings of the target domain. For example, in ‘Scott Morrison shot down all the arguments of the bushfires being a consequence of climate change.’ (BBC 2019) we can see an instance of ‘ARGUMENT (target) IS war (source)’ mapping. The underlined expression ‘shot down all the arguments’ here represents one possible lexicalisation of that conceptual mapping, others being ‘his claims are indefensible’, ‘she attacked each of my arguments’, and more.

Metaphors as semantic and cognitive phenomena are often purposefully utilised in persuasion-aimed discourses, such as political, organisational or more specific corporate communication, in order to create a specific frame, a convincing belief structure by subliminally activating our preconscious evaluation systems (Charteris-Black 2005, 2). The cross-domain pairing evident in metaphors is important in image construction, which is in turn crucial in forming and presenting desired ideas to a given public (Dobrić 2009). Speakers, lecturers, writers, and journalists utilise metaphors, along with other rhetorical ploys, to legitimise or delegitimise certain stances (Chilton 2004, 23-47). All metaphors used for a persuasive purpose on the most basic level of cognition (Dobrić and Weder 2015), both in an issue-positive and an issue-negative way, add up to what we termed



framing and create an evaluative framework whose structure represents the basis firstly for frames and then, cumulatively, for specific narratives.

The given cognitive, and by extension semantic, transparency and image-evoking quality of metaphors is what makes them prime candidates for investigating the representations of sustainability of resources within the framework of anthropogenic climate change, as delivered by and represented in the media.

### **3 Expanding the Framework to Deconstruct Sustainability Communication - A Methodological Perspective**

Having accounted for the theoretical foundations of the proposed research, the question remains of how we can identify issues and the most prevalent ‘narratives’, before going deeper and thus pinpoint frames as the interpretative packages that give meaning to the issue (Gamson and Modigliani 1989) and metaphors as communicative moments of framing processes. The innovative character of our theoretical as well as following methodological concept is that we understand metaphors as representations or building-blocks of a specific frame which, again, is the organising principle of a wider narrative. In other words, identification of metaphors unveils the related frames and, by extension, related narratives. Therefore, investigating the conceptualisation of sustainability done through metaphors is the key for reconstructing a wider understanding of the existing narratives of sustainability. A possible methodological framework for doing so is offered in the following sections.

#### **3.1 Understanding Nature by Studying Frames and Metaphors**

Framing the environment and nature is an established area of study in environmental communication research. For instance, Miyase et al. (2018) deal with the use and construction of the environment and environmental themes in different spheres such as literature, media, film, social movements, and politics. Beckmann et al. (2001) describe how news media and journalists construct and frame environmental issues (on the example of climate change). Brüggemann contributes to this research field with his studies on journalistic framing and the construction of climate change in the media (2014). Similarly, Lidström’s (2018) work on framing the rise of the sea-level is of significance for a possible study (and it is based on Nisbet and Newman’s (2015) review of framing research and understanding public discourses and debates regarding the environment). There are also a few studies focusing on metaphors of nature or metaphors in environmental communication in general. Norgaard deals with the metaphor of nature as a “fixed stock of capitalism [...] that provides a flow of services, which is insufficient for the difficulties we

are in or the task ahead” (2010, 2417), in conjunction with the concepts of ‘ecosystem services’ and ‘environmental governance’. Deigan, Semino, and Paul (2019) give an in-depth account on metaphors found in climate-related scientific discourse, focusing on research articles, educational texts, and secondary school students’ utterances. Finally, Romaine evaluated with “Greenspeak” the “role metaphorical thought plays in the scientific as well as popular discussion of key environmental issues such as global warming and loss of biodiversity” (1996, 175). Her discussion of conceptual metaphors used in environmental discourse and in how they are ideologically loaded (*ibid.*, 176) makes her article one of the core inspirations for our project. However, she mainly focuses on “clean” and “green” as synonyms for being morally good (*ibid.*, 176), following up from Harré et al. (1999), while the study-concept offered here wants to go one step further and operate with cognitive heuristics (Mio, 1997) and possible clusters (linked to the previously described framing approach) of “nature as economy”, “nature as home”, “nature as music”, “nature as living being”, “nature as miracle”, “nature as agricultural crop”, and more (Meisner 1995).

There are also the descriptions of enemy-victim metaphors, outlining the destructive relationship between humans and nature; here, ‘war’, ‘cancer’, ‘parasite’, and ‘predator’ conceptualisations can be found as used (Marshall and Toffel 2005). Keulartz dealt with metaphors for nature originating from the domains of engineering, cybernetics, art and aesthetics, medicine and health care, and geography (2007, 29ff). Another related study comes from Brown (2013), who in his analysis of NGOs and Green businesses and the corresponding framing processes finds lexical representations of a system of concern and frames for biodiversity (including ‘degradation’, ‘resources’, ‘habitats’, ‘impact’, and more). He discusses frames of ‘responsibility’, frames of ‘risk management’, and frames of ‘perception of the damage’, though he does not venture beyond the lexical level of analysis. An interesting study from Renzi et al. (2017) focuses on the problem of nuclear power in social discourses, language, and public choice, and works with three categories in which metaphors can be classified, namely ‘rebirth’ (renaissance), ‘devastation’ (apocalypse, inferno, genie, and bomb) and ‘sickness’ (addiction). Additionally, Atanasova and Koteyko’s study on conceptual metaphors points out their potential to communicate the urgency to act on climate change in two online newspapers (2015); the same was done in Grevsmühl’s study on the ‘ozone hole’ metaphors (2018). While agenda setting sits as a special research area within media studies dealing with how the media paint particular images of a given issue for the consummation by their audiences and what sources of information and streams of influence can affect the manner in which they construct the given representations, framing is about how certain aspects of an intended message (in the media) can be made more visible in order to communicate a particular issue (as,

perhaps, part of an intended agenda). Following this approach and interest in ‘how’ the message is transferred and not just in ‘what’ is communicated, we went one step further in our theoretical framework and complemented the media and communication perspective with a linguistic perspective, bringing in *conceptual metaphor theory*, as the arguably most salient sub-field of cognitive linguistics. Conceptual metaphors have only rarely been studied in the field of environmental communication and this has not been appropriately applied to sustainability communication in particular. In this respect, the following section offers a potentially useful methodology for future research in this field.

### 3.2 A Research Methodology to Understand Sustainability

Following from the previous argumentation that metaphors break up phenomena which are hard to grasp and play a major role in any attempt to frame and set an agenda (whether in political, media, or any other type of discourse), the proposition is that therefore they may also play an important role in environmental and sustainability communication. In order to explore this proposition, any appropriate research approach would have three research questions to answer at its core:

1. How is the sustainability of resources (with all of its climate change- and climate crisis-related subdomains) conceptualised in the media in general?
2. What are the possible emotions (fear, anxiety, relief, indifference, ridicule, and more) mirrored in the discovered conceptualisations?
3. What hypothetically influences the choice of conceptualisation the daily media adopt when reporting, i.e. what are the prospective sources (scientific community, governments, industry) the media could tap in terms of conceptualisation patterns employed in their own framing and agenda setting activities?

In order to answer the stated research questions, there are several practical steps that need to be undertaken. They include data collection (the compilation of the relevant corpora), metaphor (and conceptualisation pattern) identification, and frames and agenda delineation. The data collection highly depends on the focus of the study as well as the accessibility of data and is a consideration for each individual study. Therefore, we can only focus here on the overarching element of the procedure of metaphor identification. The flow of the metaphor identification is intended to follow the standard Pragglejaz Metaphor Identification Procedure (MIP) (Steen 1999a, Pragglejaz Group 2007, Steen 2007, Steen et al. 2010). After reading the entire text in question to gain a general understanding of the overarching discourse (Dobrić 2014, 145) as step 1, step 2 involves carefully reading the entire text a second time, looking for lexical items of interest (potential metaphors related to the research at hand) and accurately establishing their basic

(prototypical) and immediate (contextual) readings. The meanings of the discovered relevant lexical items (metaphors) are to be weighted in terms of their contextual readings by comparing them with the first listed general readings in a referential dictionary, which can be taken as a benchmark of basic prototypical meanings of the given expressions. In step 3, the lexical units at hand are then marked as metaphorical (in a corpus) if their immediate readings are related to their basic meanings through some form of a similarity relationship, as previously indicated. Step 4 involves the understanding of the similarity of the comparison salient in the identified metaphor. Finally, step 5 of any such analysis involves the identification of the conceptual metaphors, or rather the conceptual mappings, after the metaphorical expressions discovered in the previous steps get to be grouped by similarity of meaning. Depending on the focus of the research, both the targets and the sources in the conceptual mappings are of potential interest. Additionally, given that most conceptual mappings encompass a concrete-to-abstract transfer (Dobrić 2010, 33–35), the identification of the domains involved is additionally guided by this principle. Figure 2 below illustrates the entire procedure.

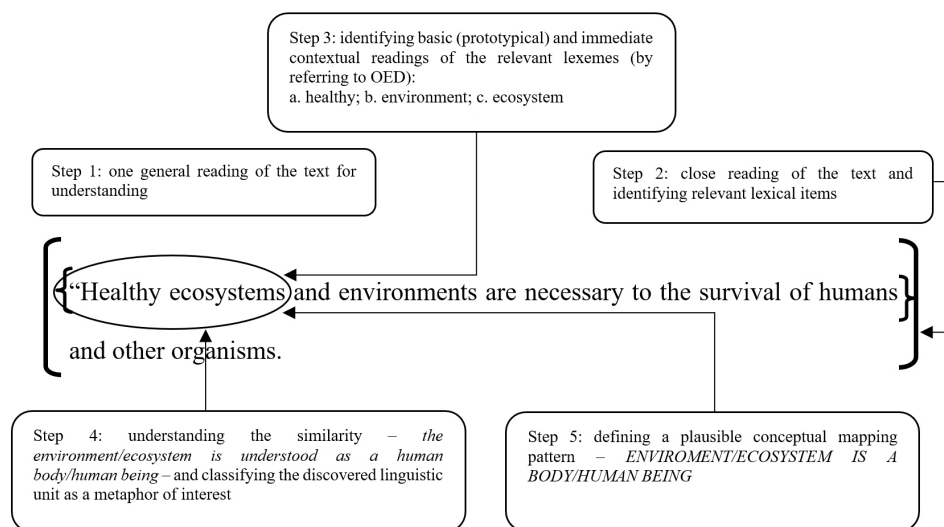


Figure 2: An example of the MIP being applied to a hypothetical example.

Once identified, the metaphors can be (electronically) marked in the corpus representing the relevant data set. Once all of the gathered texts of interest are processed in this manner, all of the marked metaphors can be extracted and their mappings once more reviewed (and, if needed, corrected). This procedure not only helps in understanding the conceptualisation patterns but also allows for a creation of an annotation scheme which can be used further on in processing any related corpora in the future. The main purpose of such a research endeavour would be, as indicated, to use the discovered conceptualisation pat-

terns for identifying any occurring frames in the processed texts and, by extension, in the related discourse.

### 3.3 Outlining Frames and Agenda(s) by Using Identified Metaphors

The identification of metaphors in texts is a crucial part of any similar study, but it is only the first step towards a full understanding of the manner in which sustainability and resources are presented and subsequently perceived. They are to represent the smallest meaningful (conceptual) parts making up frames and it is by identifying their possible groupings according to domain(s), both source and target, that frames themselves can be identified. It is a common practice to look at the imagery expressed by a number of metaphors and to try and cluster them by similarity (as for example with the STRICT FATHER vs. NURTURANT PARENT conceptualisation patterns identified in political discourse in the United States by Lakoff (2002)). It is important here to point out the criteria via which a cluster signified by similarly conceptualised metaphors is to be considered a frame:

1. semantic similarity of the lexicalisation of the metaphors at hand
2. lexical, semantic and/or conceptual similarity of the mapping domains (source or target or both); and
3. sufficient frequency of occurrence relative to the size of the corpus.

Once different clusters are identified to a sufficient degree of confidence in terms of the similarity of conceptualisation and in terms of frequency of appearance for us to claim the pattern to constitute a frame, the next step is to group frames according to similarity and reinforcement (repetition) and thusly identify agendas as wider sets of messages presented to the public. As an extra step, because of the intermedial nature of journalistic reporting, links to similar agendas (as well as frames and metaphors) in other, possibly ‘organising,’ texts could additionally be sought, including industry, government, and scientific texts (compiled into the said cross-reference corpora). The entire methodology is summed up briefly in Figure 3.

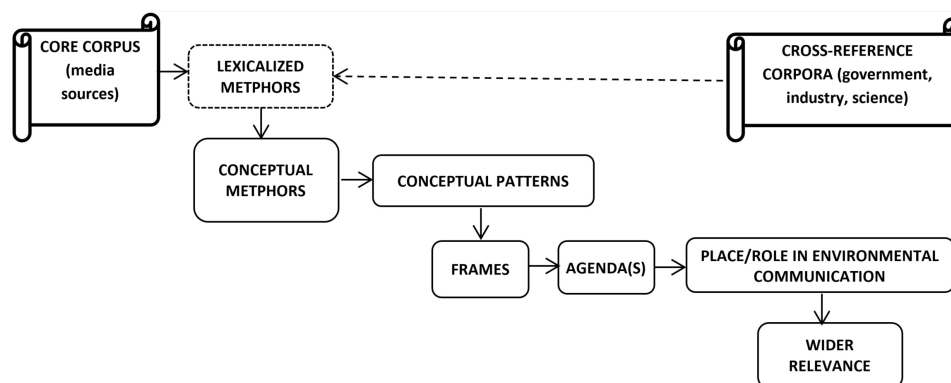


Figure 3: A graphical representation of the methodological steps planned for the project implementation, obtainable results, and the relevant sources of data.

However, embedded in a framework of narratives and framing, our concept focuses on metaphors. Likewise, the transition of metaphors from one media discourse to another is captured with the model at hand which allows for further conclusions about narratives and framing processes. In the final discussion, we will give an outlook for future studies and possible applications of the concept.

## 4 Discussion

While there is a long history of thinking about how people perceive their environment and how much this influences their knowledge and their awareness of problems (as well as their engagement), there are few and only limited studies dealing with sustainability communication, and even fewer dealing with meaning-making processes in relevant communicative contexts (e.g. corporate, political, or science communication) and specific countries or cultural settings. However, seeing that ‘concern’ is the variable that stimulates people to move from the first stage of simple recognition of the environment to information-seeking behaviour and, finally engagement (Schiff 1980), the relevant meaning-making processes are extremely important for us to understand, especially assuming that there is a strong link between media, communication, and environment on one side and language and rhetoric and discourse in public communication on the other (Peeples 2015). With the perception of narratives and framing presented here as a dynamic process of opinion-formation that is circumstantially-bound and in which the prevailing modes of presentation in elite rhetoric and news media coverage shape mass opinion (Iyengar 1991, 67), it is important to not only focus on what is being communicated, but also on the variations in how a given piece of information is being presented in the sense of organised and publicly-shaped discourse and how metaphors play a role in this

process. Being challenged by a situation where it is hard to deconstruct ‘sustainability’ as ‘masterframe’, here, in this contribution, we offer a framework to work with metaphors as the ‘smallest units’ in communication and the wider transition of metaphors in the sense of an agenda-setting process, seeing them as very specific and intentionally used tools employed to create an explicit meaning, and thus something where the process of sensemaking originally concentrates. For the application of the framework and methodological approach at hand, the data collection has to be elaborated carefully, acknowledging language barriers (metaphors might vary between different languages and time lapses (metaphors might change over time, comparisons and longitudinal studies might be fruitful), as well as it not being possible to ‘just’ translate metaphors from one language into another (see Dobrić 2011). In a wider sense, with this concept, we want to stimulate further research in this area and confirm and stimulate the interdisciplinary and transdisciplinary character of sustainability communication.

## References

- Abbott, H.P. (2009). *The Cambridge Introduction to Narrative*. 2<sup>nd</sup> ed., Cambridge: Cambridge University Press.
- Adomßent, M., J. Godemann, C. Herzig, M. Rieckmann, and D. Fischer. 2013. “Higher Education for Sustainable Development: Moving the Agenda Forward.” *Journal of Cleaner Production* 62: 1–138.
- Atanasova, Dimitrinka and Nelya Kotevko. 2017. “Metaphors in *Guardian Online* and *Mail Online*. Opinion-page Content on Climate Change: War, Religion, and Politics.” *Environmental Communication* 11 (4): 452–469. <https://doi.org/10.1080/17524032.2015.1024705>
- BBC. 2019. *Is climate change to blame for Australia’s bushfires?* <https://www.bbc.com/news/world-australia-50341210>.
- Beckmann, S. C., Christensen, A. S. and Christensen, A. G. 2001. “Myths of nature” and environmentally responsible behaviours: An exploratory study. Paper presented at the 30th European Marketing Academy Conference, Bergen/Norway.
- Brown, Mark. 2013. “A Methodology for Mapping Meanings in Text-Based Sustainability Communication.” *Sustainability* 5 (5): 2457–2479.
- Brüggemann, Michael. 2014. “Between Frame Setting and Frame Sending: How Journalists Contribute to News Frames.” *Communication Theory* 24 (1): 61–82.
- Charteris-Black, Jonathan 2005. *Politicians and Rhetoric. The Persuasive Power of Metaphor*. London: Palgrave Macmillan.
- Chilton, Paul. 2004. *Analysing Political Discourse. Theory and practice*. London: Routledge.
- Cobb, R., J.-K. Ross, and M. H. Ross. 1976. “Agenda Building as a Comparative Political Process.” *The American Political Science Review* 70 (1): 126–138. <https://doi.org/10.2307/1960328>
- De Vreese, Claes. 2005. “News Framing: Theory and Typology.” *Information Design Journal* 13 (1): 51–62.

- Dobrić, Nikola. 2009. "Metaphor Choice in Serbian Political Speech – The President and the Image of the Benevolent Father." *Political Linguistics* 30 (4): 93–99.
- Dobrić, Nikola. 2010. "Theory of Names and Cognitive Linguistics: The Case of the Metaphor." *Filozofija i društvo*, 21 (1): 135–147.
- Dobrić, Nikola. 2011. "The Cognitive Approach to Translating Metaphors Revisited – the Case of Pure, Clear and Clean vs. Cist and Jasan." *AAA, Arbeiten aus Anglistik und Amerikanistik* 36 (2): 99–117.
- Dobrić, Nikola. 2014. On Some Problems of Meaning – Polysemy between Sense Enumeration and Core Meaning Paradigms. *Filozofija i društvo*, 25 (4): 146–163.
- Dobrić, Nikola and Franziska Weder. 2015. "Media conceptualizing illnesses – the case of the flu" *Journal of Media & Cultural Studies* 30 (1): 126–142.
- Eisenstein C. (2013) *The more beautiful world our hearts know is possible*. North Atlantic Books, Berkeley.
- Evans, Vyvyan. 2004. *The Structure of Time: Language, Meaning and temporal cognition*. Amsterdam: John Benjamins Publishing.
- Evans, Vyvyan. 2005. "The meaning of time: Polysemy, the lexicon and conceptual structure." *Journal of Linguistics* 41 (1): 33–75.
- Frank, A. K. (2017). What is the story with sustainability? A narrative analysis of diverse and contested understandings. *J Environ Stud Sci* 7, 310–323.
- Gamson, William A. and Andre Modigliani. 1989. "Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach." *American Journal of Sociology* 95 (1): 1–37.
- Gerhards, Jürgen and Dieter Rucht. 1992. "Mesomobilization: Organizing and Framing in Two Protest Campaigns in West Germany." *American Journal of Sociology* 98 (3): 555–596.
- Gibbs Jr., Raymond W. 2006. "Metaphor Interpretation as Embodied Simulation." *Mind & Language* 21 (3): 434–458.
- Gibbs Jr., Raymond W., ed. 2008. *The Cambridge Handbook of Metaphor and Thought*. Cambridge: Cambridge University Press.
- Gitlin, Todd. 1980. *The Whole World is Watching. Mass Media in the Making and Unmaking of the New Left*. Oakland: University of California Press.
- Glucksberg, Sam and Catrinel Haught. 2006. "Can Florida become like the Next Florida? When Metaphoric Comparisons Fail." *Psychological Science* 17 (11): 935–938.
- Glucksberg, Sam. 2008. "How Metaphors Create Categories – Quickly." In *The Cambridge Handbook of Metaphor and Thought*, edited by R.W. Gibbs Jr, 67–83. Cambridge: Cambridge University Press.
- Grevsmühl, Sebastian V. 2017. "Revisiting the 'Ozone Hole' Metaphor: From Observational Window to Global Environmental Threat." *Environmental Communication* 12 (1): 1–13.
- Grković-Mejdžor, Jasmina. 2008. "O kognitivnim osnovama semanticke promene." *Srpski jezik u svetlu savremenih lingvističkih teorija*: 49–63.
- Harré, Rom and Luk Van Langenhove eds. 1999. *Positioning Theory: Moral contexts of Intentional action*. Oxford: Blackwell.



- Herrick, Charles N., & Joanna L. Pratt. 2013. Communication and the Narrative Basis of Sustainability: Observations from the Municipal Water Sector. *Sustainability* 2013, 5, 4428–4443.
- Iyengar, Shanto. 1991. *Is Anyone Responsible? How Television Frames Political Issues*. Chicago: The University of Chicago Press.
- Keulartz, Josef. 2007. "Using Metaphors in Restoring Nature." *Nature and Culture* 2 (1): 27–48.
- Kövecses, Zoltán. 2002. *Metaphor: a Practical Introduction*. Oxford: Oxford University Press.
- Kövecses, Zoltán. 2005. *Metaphor in Culture. Universality and Variation*. Cambridge: Cambridge University Press.
- Krainer, Larissa and Franziska Weder. 2011. "Editorial" *Medien Journal* 35 (1): 2–3.
- Lacy, Stephen and David C. Coulson. 2000. "Comparative Case Study: Newspaper Source Use on the Environmental Beat" *Newspaper Research Journal* 21 (1): 13–25.
- Lakoff, George and Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.
- Lakoff, George. 1993. "The Contemporary Theory of Metaphor." In *Metaphor and Thought*, edited by Andrew Ortony, 202–251. Cambridge: Cambridge University Press.
- Lakoff, George. 2002. *Moral Politics. How Liberals and Conservatives Think*. 2nd ed. Chicago and London: The University of Chicago Press.
- Lang, Kurt and Gladys Engel Lang. 1959. "The Mass Media and Voting." In *American Voting Behavior*, edited by Eugene Burdick and Arthur Brodbeck, 217–235. Glencoe, III.: The Free Press.
- Lidström, Susanna. 2018. "Sea-Level Rise in Public Science Writing: History, Science and Reductionism." *Environmental Communication* 12 (1): 15–27. New Jersey: Blackwell Publishers.
- Marshall, Julian D. and Michael W. Toffel. 2005. "Framing the Elusive Concept of Sustainability: A Sustainability Hierarchy." *Environmental Science & Technology* 39 (3): 673–682.
- McCombs, M. E. and D. L. Shaw. 1972. "The Agenda-Setting Function of Mass Media." *The Public Opinion Quarterly* 36 (2): 176–187.
- McCombs, M. E., J. P. Llamas, E. Lopez-Escobar, and F. Rey. 1997. "Candidate Images in Spanish Elections: Second-Level Agenda-Setting Effects." *Journalism & Mass Communication Quarterly* 74 (4): 703–717.
- McCombs, M. E. 2004. *Setting the Agenda: The Mass Media and Public Opinion*.
- McCombs, M. E. 2005. "A Look at Agenda-setting: past, present and future." *Journalism Studies* 6 (4): 543–557.
- Meisner, Mark S. 1995. "Metaphors of Nature: Old Vinegar in New Bottles?" *The Trumpeter* 12 (1): 11–18.
- Mio, Jeffrey Scott. 1997. "Metaphor and Politics." *Metaphor and Symbol* 12 (2): 113–133. [http://dx.doi.org/10.1207/s15327868ms1202\\_2](http://dx.doi.org/10.1207/s15327868ms1202_2)
- Moser, Susanne C. 2010. "Communicating climate change: History, challenges, process and future directions." *Wiley Interdisciplinary Reviews: Climate Change* 1 (1): 31–53.
- Nerlich, B., N. Koteyko, and B. Brown. 2010. "Theory and language of climate change communication." *Wiley Interdisciplinary Reviews: Climate Change* 1 (1): 97–110.

- Newig, Jens. 2011. "Climate Change as an Element of Sustainability Communication." In *Sustainability Communication*, edited by Jasmin Godemann and Gerd Michelsen, 119–128. Berlin: Springer.
- Nisbet, Matthew C. and Todd P. Newman. 2015. "Framing, The Media, and Environmental Communication." In *The Routledge Handbook of Environment and Communication*, edited by Anders Hansen and Robert Cox, 235–338. New York: Routledge.
- Norgaard, Richard B. 2010. "Ecosystem Services: From Eye-Opening Metaphor to Complexity Blindness." *Ecological Economics* 69 (6): 1219–1227.
- Ortony, Andrew, ed. 1993. *Metaphor and Thought*. Cambridge: Cambridge University Press.
- Oxford English Dictionary. 2019. "Home."
- Parlour, J. W. and S. Schatzow. 1978. "The mass media and public concern for environmental problems in Canada, 1960–1972." *International Journal of Environmental Studies* 13 (1): 9–17.
- Peebles, Jennifer. 2015. "Discourse/rhetorical analysis approaches to environment, media and communication." In *Handbook of Environment and Communication*, edited by Anders Hansen and Robert Cox, 39–48. New York: Routledge.
- Pilyarchuk, Kateryna and Alexander Onysko. 2018. "Conceptual Metaphors in Donald Trump's Political Speeches: Framing his Topics and (Self-)Constructing his Persona." *Colloquium: New Philologies* 3 (2): 98–156.
- Pragglejaz Group. 2007. "MIP: a method for identifying metaphorically used words in discourse." *Metaphor and Symbol* 22 (1): 1–39.
- Reese, Stephen. 2001. "A Bridging Model for Media Research." In *Framing Public Life. Perspectives on Media and Our Understanding of the Social World*, edited by Stephen Reese, Oscar Gandy, Jr., and August Grant, 7–31. Mahwah: Lawrence Erlbaum.
- Renzi, B. G., M. Cotton, G. Napolitano, and R. Barkemeyer. 2017. "Rebirth, Devastation and Sickness: Analyzing the Role Metaphor in Media Discourses of Nuclear Power." *Environmental Communication* 11 (5): 624–640.
- Roberts, Marilyn and Maxwell E. McCombs. 1994. "Agenda setting and political advertising: Origins of the news agenda." *Political Communication* 11 (3): 249–262.
- Romaine, Suzanne. 1996. "War and Peace in the Global Greenhouse: Metaphors We Die By." *Metaphor and Symbolic Activity* 11 (3): 175–194.
- Rössler, Patrick. 1997. *Agenda-Setting. Theoretische Annahmen und empirische Evidenzen einer Medienwirkungshypothese*. Wiesbaden: Verlag für Sozialwissenschaften.
- Sahinidou, Ioanna (2016) The roots of the ecological crisis and the way out: 1 creation out of 'no thing' god being 'no thing'. *Feminist Theology: The Journal of the Britain & Ireland School of Feminist Theology* 24(3):291–298. <https://doi.org/10.1177/0966735015627971>
- Schiff, William 1980. *Perception: An applied approach*. Boston: Houghton Mifflin Harcourt.
- Shah, A. A., M. C. Giddings, J. B. Parvaz, and R. F. Gesteland. 2002. "Computational identification of putative programmed translational frameshift sites." *Bioinformatics* 18 (8): 1046–1053.

- Snow, David and Robert Benford. 1992. "Master Frames and Cycles of Protest." In *Frontiers in social movement theory*, edited by A.D. Morris and C. McClurg Mueller, 133–155. London: Yale University Press.
- Steen, Gerard J. 1999. "From linguistic to conceptual metaphor in five steps." In *Metaphor in Cognitive Linguistics: Selected papers from the 5th International Cognitive Linguistics Conference*, edited by Raymond W Gibbs and Gerard J. Steen, 57–77. Amsterdam: John Benjamins.
- Steen, Gerard J. 2007. *Finding Metaphor in Grammar and Usage: a methodological analysis of theory and research*. Amsterdam: John Benjamins.
- Steen, Gerard J., Letti Dorst, J. Berenike Herrmann, and Anna Kaal. 2010. *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. Amsterdam: John Benjamins.
- Sweetser, Kaye D. and Charles W. Brown. 2008. "Information subsidies and agenda-building during the Israel-Lebanon crisis" *Public Relations Review* 34 (4): 359–366.
- van der Leeuw, Sander (2019) The role of narratives in human-environmental relations: an essay on elaborating win-win solutions to climate change and sustainability. *Climatic Chang*: 1–11.
- Weaver, David H. 2007. "Thoughts on Agenda Setting, Framing, and Priming." *Journal of Communication* 57 (1): 142–147.
- Weder, Franziska. 2012. "'Verantwortung' als trendige Referenz der Wirtschaftsberichterstattung oder: Der fehlende öffentliche Diskurs über Corporate Social Responsibility." *uwf Umwelt-WirtschaftsForum* 19 (3–4): 185–192.
- Ziemann, Andreas. 2011. *Medienkultur und Gesellschaftsstruktur. Soziologische Analysen*. Wiesbaden: Springer SV.

# Mythos versus Utopie

## Zur Bedeutung von Kindheit bei Bruno Schulz und Theodor W. Adorno

Karoline THAIDIGSMANN

Universität Heidelberg (Germany)

### Abstract

This paper examines the astonishing parallels in the depictions of childhood experience in the works of Polish writer Bruno Schulz and German philosopher Theodor W. Adorno, which so far have not received any attention in research. For both authors, the experience of the child—be it autobiographically remembered as in Adorno, or fantastically transformed as in Schulz—bears the characteristics of transcendental experience manifest in everyday life. As similar as the childhood images and sce-

nes in Schulz and Adorno might be, their significance points in two different directions. While for Schulz, the experience of the child leads back to an original meaning of things and thus to myth, for Adorno, it points to the conditions of a new order and thus to utopia.

Keywords: *Bruno Schulz, Theodor W. Adorno, Childhood, Myth, Utopia*

(c) Karoline Thaidigsmann; karoline.thaidigsmann@slav.uni-heidelberg.de

Colloquium: New Philologies, Volume 6, Issue 2 (2021)

doi: 10.23963/cnp.2021.6.2.1

Stable URL: <https://colloquium.aau.at/index.php/Colloquium/article/view/144>

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

1938 flieht der deutsche Philosoph Theodor W. Adorno vor den Nationalsozialisten in die USA. Auf die Frage, weshalb er 1949 nach Westdeutschland zurückgekehrt sei, antwortet er: „Ich wollte einfach dorthin zurück, wo ich meine Kindheit hatte, am Ende aus dem Gefühl, daß, was man im Leben realisiert, wenig anderes ist als der Versuch, die Kindheit verwandelnd einzuholen“ (Adorno 1986, 395). Adornos Antwort klingt wie ein Echo auf das poetische Credo des polnischen Schriftstellers Bruno Schulz, der 1936 in einem Brief an den Literaturkritiker Andrzej Pleśniewicz erklärt, dass

die Art von Kunst, wie sie mir am Herzen liegt, eben ein Rückzug, wieder-gekehrte Kindheit ist. Wenn es möglich wäre, die Entwicklung rückläufig zu machen, durch einen Umweg die Kindheit wiederholt einzufangen [...] – dann wäre das die Erfüllung der „genialen Epoche“, der „messianischen Zeiten“.<sup>1</sup> (Schulz 2000a, 107)

Für Adorno wie für Schulz ist die Kindheit von elementarer Bedeutung für ihre philosophischen respektive literarischen Konzeptionen. Geht man der Bedeutung von Kindheit in Adornos und Schulz' Werk nach, so verblüffen die Gemeinsamkeiten ihrer Gedanken zum Potenzial kindlicher Erfahrung als spezifischer Erkenntnisform sowie die Ähnlichkeit der Kindheitsszenen, anhand derer sie dieses Potenzial auszuloten beziehungsweise zu evozieren suchen. Die Bedeutung dieses Potenzials legen der gesellschaftskritische Denker und der schöpferische Dichter allerdings in einander geradezu entgegengesetzter Perspektive aus. Während Kindheit für Schulz einen engen Bezug zum Mythos hat, verbindet sie sich bei Adorno mit dem Utopischen.

In der Forschung haben die Gemeinsamkeiten in der Darstellung und der Nutzung des Erkenntnispotenzials von Kindheitserfahrung bei Schulz und Adorno bislang keine Aufmerksamkeit erfahren. Die folgende Untersuchung möchte hier eine Lücke schließen. Im Zentrum des Vergleichs steht die Bedeutung von Kindheit als unreglementierte Erfahrung von Wirklichkeit. Als solche hat Kindheit sowohl bei Schulz als auch bei Adorno die Dimension einer Transzendenz Erfahrung im Alltäglichen, in der bei ersterem ein Nachglanz des Mythischen und bei letzterem ein Vorschein des Utopischen enthalten ist. Dabei stellt sich die Frage, wie weit die Perspektiven Mythos und Utopie auseinanderliegen. Auf der einen Seite weisen bereits die eingangs zitierten Aussagen darauf hin, dass die beiden Perspektiven möglicherweise nicht so stark differieren, wie es scheint. Auf der anderen Seite erlaubt die Spannung zwischen ihnen eine Konturierung des jeweils Spezifischen von Adornos philosophischer und Schulz' poetischer Weltsicht, trotz,

<sup>1</sup> „[T]en rodzaj sztuki, jaki mi leży na sercu, jest właśnie regresją, jest powrotnym dzieciństwem. Gdyby można było uwstecznić rozwój, osiągnąć jakąś określną drogą powtórnie dzieciństwo [...] – to byłoby to ziszczeniem ‚genialnej epoki‘, ‚czasów mesjaszowych‘ [...]“ (Schulz 1989, 424).

wenn nicht gerade wegen der ganz verschiedenen Zusammenhänge, in denen sie dachten, schrieben und lebten. Adorno und Schulz sind sich nie begegnet, auch rezipierten sie einander meiner Kenntnis nach nicht.

## 1 Kindheit – Biografischer Hintergrund und literarische Vergegenwärtigung

Sowohl für Adorno als auch für Schulz bildet die Welt ihrer eigenen Kindheit und die Aura, die diese in der Rückschau umgibt, den jeweiligen Ausgangspunkt ihres Nachdenkens über das Potenzial kindlicher Erfahrung. Während Schulz (1892–1942) in der galizischen Stadt Drohobycz in das Milieu jüdischer Kleinkaufleute hineingeboren wird, wächst der elf Jahre jüngere Adorno (1903–1969) im gehobenen Bürgertum in Frankfurt am Main auf. Sein Vater ist wohlhabender Weinhändler, die Mutter eine ehemalige Opernsängerin. Beide Kinder zeichnen sich durch eine Affinität zur Kunst aus. Bei Schulz wird früh, noch bevor er zu schreiben beginnt, ein zeichnerisches Talent sichtbar, Adorno sticht durch seine musische Begabung hervor. Im Werk beider Autoren werden sich später Spuren des jüdischen Erbes niederschlagen, auch wenn es in den Familien, in denen sie aufwachsen, keine zentrale Rolle mehr spielt. Schulz' Eltern pflegen die jüdische Religion nur mehr der äußeren Form nach, in Adornos Familie klingt der Bezug zum Judentum im Familiennamen Wiesengrund nach. Trotz der geographischen und kulturellen Unterschiede, in denen sie aufwachsen, und trotz ihres Altersunterschieds, erfahren Schulz und Adorno beide die Zeit ihrer Kindheit und Jugend nicht nur als eine persönlich bedeutungsvolle Zeit, sondern auch als eine Zeit der durch Technisierung und Kommerzialisierung bedingten ökonomischen und sozialen Umbrüche. Schulz erlebt diesen Wandel als existentielle Bedrohung für seine Familie. Adorno nimmt die Veränderungen als Kind eher aus beobachtender Distanz wahr.<sup>2</sup>

Die Art und Weise, in der Kindheit später in Schulz' und Adornos Werken präsent wird, unterscheidet sich grundsätzlich voneinander. Während Schulz' gesamtem poetischem Kosmos die Idee der Kindheit zugrunde liegt und dessen kreativer Motor ist, werfen Adornos vereinzelt über sein Werk verstreute Reflexionen über die Kindheit eher wie beiläufig ein Licht auf zentrale Motive seines Denkens. Schulz versucht in seinem Realität mit Phantastik verbindenden Erzählwerk, das vor allem aus den Erzählbänden *Sklepy cynamonowe* (1933; *Die Zimtläden*) und *Sanatorium pod klepsydrą* (1937; *Sanatorium zur Sanduhr*) besteht, weniger Kindheit darzustellen, als vielmehr die Erfahrungsweise des

<sup>2</sup> Zu den biografischen Angaben über Adorno siehe Jäger (2009, 9–25), über Schulz siehe Ficowski (2003, v. a. 33–43) sowie die Nachworte von Jerzy Jarzębski in Schulz (1989) und Mikołaj Dutsch in Schulz (2000b).

Kindes mit poetischen Mitteln erlebbar zu machen. Die Leserschaft soll in eine literarisch geschaffene Welt hineingezogen werden, die nach den Regeln kindlicher Wahrnehmung funktioniert. Adorno hingegen geht es nicht um eine solch unmittelbare Rückführung seiner Leser in die Erfahrungswelt des Kindes. Nicht das (Wieder-)Erleben der Kindheit, sondern die Reflexion der Erfahrungsweise des Kindes steht für ihn im Zentrum. „Besonders produktiv war Adorno oft dann“, bemerkt sein Biograph Lorenz Jäger, „wenn er die Erinnerungen an die Kindheit abrufen und theoretisch einsetzen konnte“ (Jäger 2009, 24).<sup>3</sup> Trotz Adornos primär theoretischem Interesse an Kindheitserfahrungen fällt auf, dass sich Reminiszenzen an die Kindheit gerade in denjenigen seiner Texte finden, die einen literarischen Charakter haben. Dabei handelt es sich vor allem um die Aphorismensammlung *Minima Moralia* (1951), die er während und kurz nach dem Zweiten Weltkrieg schrieb, sowie um sechzehn kurze Prosaskizzen, die 1966 unter dem Titel *Amorbach* in der *Süddeutschen Zeitung* erschienen. Adorno mag, was den Skizzencharakter seiner Kindheitserinnerungen anbelangt, von Walter Benjamins in den 1930er-Jahren verfassten autobiographischen Kindheitsminiaturen *Kindheit um 1900* angeregt worden sein, die inzwischen auch in der Schulz-Forschung auf Interesse gestoßen sind.<sup>4</sup> Adorno hatte die Manuskriptseiten des Freundes 1950 zu einer ersten publizierten Buchausgabe zusammengefügt und mit einem Nachwort versehen.

## 2 Kindheit – Transzendenzerfahrungen im Alltäglichen

Das Potenzial, das Kindheit für Schulz zur „genialen Epoche“ werden lässt, besteht in der spezifischen Form der Wirklichkeitserfahrung des Kindes. In der gleichnamigen Erzählung *Genialna epoka* (*Die geniale Epoche*) ruft der kindliche Erzähler – in seiner Phantasie angeregt durch die Farbspiele, mit denen das Ofenfeuer und das Licht der letzten Wintertage die Wohnung verwandeln – seiner Familie entgegen:

„Seht ihr [...], immer habe ich euch gesagt, daß alles gehemmt ist, zugemauert vor Langeweile, nicht erlöst. Und jetzt schaut, was für eine Überschwemmung, was für ein Blühen, was für eine Seligkeit...“ [...] Doch [...]

<sup>3</sup> Nach Philipsen (2007, 108) ist die „Privilegierung von Kindheitserfahrung“ in allen Etappen von Adornos Werk präsent, gewinnt seine besondere Bedeutung jedoch v. a. in Adornos späten Arbeiten.

<sup>4</sup> 2016 erschien eine vergleichende Studie zu Schulz' Erzählungen und Benjamins *Kindheit um 1900* (Czabanowska-Wróbel 2016). Dort auch ein knapper Überblick über weitere Studien, die Bezüge zwischen Schulz und Benjamin herstellen. Benjamin hat für die Schulz-Forschung vor allem im Kontext einer Entdeckung der jüdischen Tradition in Schulz' Werk Bedeutung gewonnen.

alle standen ratlos da und blickten sich an und versteckten sich hinter den Rücken der Nachbarn.<sup>5</sup> (Schulz 2000b, 120–121)

An den Reaktionen der Umstehenden zeigt sich: Es ist nicht so sehr die Kraft des Lichts, die dem gewöhnlichen Zimmer eine veränderte Gestalt verleiht, als vielmehr die Bereitschaft des Kindes, die durch Gewohnheit bestimmte Wahrnehmung seiner Umwelt aufzugeben.<sup>6</sup> Der streng reglementierte und nach praktischen Gesichtspunkten geordnete Alltag der Erwachsenen begrenzt die Wahrnehmung der Wirklichkeit nach Kriterien der Zweckhaftigkeit. „Luftdicht abgeschlossen“ (Schulz 2000b, 128)<sup>7</sup> bietet diese Wirklichkeit keine Freiräume, um die Dinge aus der engen zweckorientierten Wahrnehmung, in der sie „eingemauert“ sind, herauszulösen. Mit Besen und Wischtuch ausgerüstet steht in Schulz' Erzählungen das Hausmädchen Adela allegorisch für solch praktische Lebensorientierung und weist mit „eiserne[r] Disziplin“ (ibid., 81)<sup>8</sup> jedem Ding seinen unbedingten Platz zu. Eine Wirklichkeit jedoch, in der den Dingen kein über ihren praktischen Nutzen hinausreichender Eigenwert zugesprochen wird, lässt diese austauschbar und die Wirklichkeit damit „grau“ und „langweilig“ werden, zwei Begriffe, die Schulz' Prosa leitmotivisch durchziehen. Das Kind hingegen nimmt seine Umwelt nicht nach praktischen Gesichtspunkten wahr, vielmehr wendet es sich dem einzelnen zu und macht es zum Sprungbrett seiner Phantasie. Diese Fähigkeit, in seiner konkreten Lebenswelt unter der Oberfläche ihrer Alltäglichkeit transzendente Erfahrungen zu machen, lassen das Kind, nach Schulz, zum eigentlichen Wesen der Dinge vordringen. Die „Buntheit der genialen Epoche“ (ibid., 80) stellt sich der grauen Langeweile der praktischen Alltagswelt als eine Zeit der „Sonntage und Feiertage“ (ibid.) entgegen.<sup>9</sup>

Schulz löst in seiner Prosa das Feiertägliche von seiner äußeren ritualisierten Form, hält aber die Bedeutung religiöser Feste und Feiertage als aus der Alltagswirklichkeit herausgehobene Zeit fest, die ein lebensförderndes Zu-Sich-Selbst-Kommen der Dinge er-

<sup>5</sup> In Fällen, in denen die publizierten deutschen Übersetzungen stark vom Original abweichen, wurden sie von mir verändert. „– Widzicie – krzyczałem [...] – zawsze mówiłem wam, że wszystko jest zatamowane, zamurwane nudą, nie wyzwolone. A teraz patrzcie, co za wylew, co za rozkwit wszystkiego, co za błogość... [...]. Ale [...] stali bezradni i oglądali się za siebie, cofali za plecy sąsiadów“ (Schulz 1989, 123).

<sup>6</sup> Zur Spontanität des Erlebens als zentralem Merkmal des Kindseins bei Schulz siehe auch Jarzębski (2003, 10).

<sup>7</sup> „zamknięte na głucho“ (Schulz 1989, 131). Aus der Erzählung *Genialna epoka*.

<sup>8</sup> „wszystkie sprzęty poddały się żelaznej dyscyplinie, jaką Adela roztoczyła nad tym pokojem“ (Schulz 1989, 82). Aus der Erzählung *Karakony* (*Die Küchenschaben*).

<sup>9</sup> „Było to w okresie szarych dni, które nastąpiły po świetnej kolorowości genialnej epoki [...]. Były to [...] ciężkie tygodnie bez niedziel i świąt“ (Schulz 1989, 81). Aus der Erzählung *Karakony*. In dieser und anderen Erzählungen wird deutlich, dass in Schulz' Prosawerk häufig auch die Vaterfigur in einer gleichsam kindlichen „genialen Epoche“ verbleibt, allerdings mit durchaus ambivalenten Konsequenzen.



möglichst.<sup>10</sup> Doch gibt er den Wert der Ritualisierung der Feiertage nicht ganz preis. Festgelegte Feiertage bieten zumindest äußere Bedingungen, die die alltägliche Ordnung aufheben können und so eine veränderte Wahrnehmung der Wirklichkeit begünstigen; eine Funktion, die für Schulz auch die Ferienzeit erfüllt. Nicht zufällig bilden daher religiöse – jüdische wie christliche – Feiertage und Ferienzeiten in vielen Fällen den Hintergrund seiner Erzählungen. Der Abschied von der dem praktischen Leben enthobenen Kindheit und das Ende der Ferien fallen metaphorisch zusammen:

Und wir packen ebenfalls die Koffer. Ich bin fünfzehn Jahre alt und stehe den Pflichten des praktischen Lebens völlig unerfahren gegenüber. Weil noch eine Stunde bis zur Abfahrt ist, laufe ich noch einmal hinaus, um von der Sommerfrische Abschied zu nehmen [...], zu sehen, was man mitnehmen kann und was man schon für immer [...] zurücklassen muß.<sup>11</sup> (ibid., 329)

Ganz ähnlich wie für Schulz liegt für Adorno das Potenzial der Kindheit in der Erfahrungsoffenheit des Kindes, die es diesem ermöglicht, Wirklichkeit jenseits der von praktischen Zwecken reglementierten Lebensordnung wahrzunehmen. Die Suspendierung der Alltagsordnung – auch bei Adorno in der religiösen Kategorie des Feiertags gedacht, zugleich vom Religiösen gelöst und an die Erfahrung des Kindes gebunden – setzt erst das eigentliche Leben wieder in sein Recht, wie eine der schönsten Kindheitsreflexionen Adornos unter dem Titel „Heliotrop“ deutlich macht:

Dem, zu dessen Eltern Logierbesuch kommt, schlägt das Herz mit größerer Erwartung als je vor Weihnachten. Sie gilt nicht Geschenken, sondern dem verwandelten Leben. [...] Die Koffer [...] sind Truhen, in denen die Edelsteine Aladins und Ali Babas, eingehüllt in kostbare Gewebe, die Kimonos des Logierbesuchs, aus den Karawansereien der Schweiz und Südtirols in Schlafwagensänften hergeschleppt werden zur gesättigten Betrachtung. [...] Mit der Ordnung des Tages – vielleicht darf am folgenden die Schule versäumt werden – sind auch die Grenzen zwischen den Generationen suspendiert, und die wahre Promiskuität ahnt, wer um elf Uhr immer noch

<sup>10</sup> Zur biblischen Bedeutung von Festen und Feiertagen vgl. Koch u.a. (2000, 145–147, 438–439). Die Lösung des Feiertäglichen von seinem religiösen Kontext kann als ein Aspekt des für Schulz' Werk charakteristischen und von der Schulz-Forschung dokumentierten Spiels mit Sakralem und Profanem betrachtet werden. Zu Profanierungen bei Schulz siehe z. B. Prokopczyk (1999).

<sup>11</sup> „I my także pakujemy kufry. Mam piętnaście lat i jestem całkiem nie obciążony obowiązkami praktyki życiowej. Ponieważ jest jeszcze godzina do wyjazdu, wybiegam jeszcze raz, pożegnać letnisko, [...] zobaczyć, co można zabrać ze sobą, a co już trzeba na zawsze zostawić w tym mieście [...]“ (Schulz 1989, 320). Aus dem Textfragment *Jesień* (*Der Herbst*).

nicht ins Bett geschickt wird. Der eine Besuch weiht den Donnerstag zum Fest [...]. Liebe zählt die Stunden bis zu jener, da der Logierbesuch über die Schwelle tritt und das verfärbte Leben wieder herstellt durch ein Unmerkliches. (Adorno 2003a, 201–202)

Was „dem Leben den Zauber in späteren Jahren“ nimmt, so Adorno (ibid., 259) im Prosastück *Kaufmannsladen*, ist die Beschränkung von Dingen und Handlungen auf ihren nur funktionalen Gebrauchswert. Nicht mehr „das Licht der eigenen Bestimmung“ (ibid., 260), sondern nur ihr abstrakter Tauschwert sind von Bedeutung. Damit aber werden sie einförmig: die Wirklichkeit verfärbt sich, alles wird in gleichmacherisches „Grau“ getaucht (ibid.). Für Adorno ist diese Reduktion Merkmal der modernen „Warenwelt“ (ibid.). Mit dieser Kritik an der Warenform verknüpft sich für Adorno als ein grundsätzliches Problem der Moderne, dass der Mensch seine Wirklichkeit begrifflich ordnend wahrnimmt. Indem das Denken Kategorien und Begriffe bildet, denen die Einzelerfahrungen auf der Grundlage ihrer Gemeinsamkeiten untergeordnet werden, reduziert es unsere Erfahrung und Erkenntnis der Wirklichkeit. Dasjenige, was an den verschiedenen Erfahrungen besonders, oder wie es bei Adorno heißt, „nicht-identisch“ ist, geht dabei verloren. Was Adorno daher einfordert, ist das Festhalten an unreglementierter Erfahrung. Eine solche Erfahrungsweise findet er im „zwecklose[n] Tun“ (ibid.) des kindlichen Spiels:

Gerade indem es die Sachen, mit denen es hantiert, ihrer vermittelten Nützlichkeit entäußert, sucht es [das Kind; KT] im Umgang mit ihnen zu erretten, womit sie den Menschen gut und nicht dem Tauschverhältnis zu Willen sind, das Menschen und Sachen gleichermaßen deformiert. (ibid.)

Auf sprachlicher Ebene steht für die aus der unreduzierten Erfahrung entspringende Erkenntnis der Name. Während die begriffliche Sprache, nach Adorno, „für das Allgemeine an den Einzeldingen [steht], [stehen] Namen für das je Besondere an ihnen“ (Schwepenhäuser 1996, 67).<sup>12</sup> Der Name gibt jedem Ding seine Identität. In Schulz' Erzählung *Genialna epoka* kommt die Aufgabe identitätsstiftender Namensgebung dem von seiner kindlichen Phantasie bestürmten Ich-Erzähler zu:

Mit jeder Stunde drangen die Visionen zahlreicher auf mich ein, [...] bis [sich] eines schönen Tages [...] die ganze Gegend auflöste in Wanderzüge [...] – in endlose Pilgerscharen von Bestien und Tieren. [...] Warteten sie, um von mir den Namen und die Lösung ihres Rätsels zu erfahren, das sie

<sup>12</sup> Vgl. dazu Adorno (1973, 61). Zur Bedeutung des Namens bei Adorno siehe auch Tiedemann (1993).

nicht verstanden? Fragten sie mich nach ihren Namen, um in diese einzugehen und sie mit ihrem Wesen zu erfüllen? [...] Sie wichen mit gesenkten Köpfen zurück, [...] und kehrten, aufgelöst in ein namenloses Chaos, in die Rumpelkammer der Formen zurück.<sup>13</sup> (Schulz 2000b, 122–123)

Die Erfahrungen des Kindes sind an seine konkrete und besondere Lebensumwelt gebunden. Indem diese seine Erfahrungen prägt, wird sie zum Archetypen bzw. Prototypen von Erfahrung, oder, wie Adorno sagt, erfahrungsstiftend. Sowohl bei Schulz als auch bei Adorno ist die spezifische Lebenswelt des Kindes als Raum zu verstehen, in dem sich individuelle Erfahrung zu universeller Bedeutung hin öffnet. Schulz' phantastisch-literarischer Kosmos entfaltet sich in und um seine galizische Heimatstadt Drohobycz, die in den Erzählungen nicht namentlich genannt wird, in ihren geographischen, sozio-graphischen und ökonomischen Bedingungen jedoch durchscheint.<sup>14</sup>

Wie es ausdrücken? Während sich andere Städte wirtschaftlich entwickelt haben, in statistischen Ziffern gewachsen sind [...], hat unsere Stadt an Wesenhaftigkeit zugenommen. Hier geschieht nichts umsonst, ereignet sich nichts ohne tiefen Sinn und ohne Vorbedacht. Hier sind die Ereignisse keine ephemeren Phantome an der Oberfläche, hier wurzeln sie in der Tiefe der Dinge und streben nach Wesentlichem. Hier entscheidet sich jeden Augenblick etwas, beispielhaft und für alle Zeiten.<sup>15</sup> (Schulz 2000b, 335)

Für Adorno gewinnt eine solch archetypische Bedeutung das im Odenwald gelegene Provinzstädtchen Amorbach, in dem er die meisten Sommer seiner Kindheit verbrachte:

der Unterschied zwischen Landschaften und Gegenden, welche über die Bilderwelt einer Kindheit entscheiden, [ist] vermutlich gar nicht so groß. [...]. Aber damit dies Allgemeine, das Authentische [...] sich bildet, muß man hingerissen sein an dem einen Ort, ohne aufs Allgemeine zu spielen.

<sup>13</sup> „Z godziny na godzinę coraz tłumniej napływały wizje, [...] aż pewnego dnia [...] kraj cały rozgałęził się wędrówkami [...] – nieskończonymi pielgrzymkami bestyj i zwierząt. [...] Czy czekały, żebym je nazwał, rozwiązał ich zagadkę, której nie rozumiały? Czy pytały mnie o swe imię, ażeby w nie wejść i wypełnić je swoją istotą? [...] Wycofywały się tyłem, pochylając głowę i patrząc spode łba, i gubiły się same w sobie, wracały, rozwiązując się w bezimienny chaos, w rupieciarnię form“ (Schulz 1989, 124–125).

<sup>14</sup> Zu den Realien von Drohobycz und Umgebung, die in Schulz' Erzählungen durchscheinen, siehe Budurowycz (1986).

<sup>15</sup> „Jak to wyrazić? Gdy inne miasta rozwinęły się w ekonomikę, wyrosły w cyfry statystyczne [...] – miasto nasze zstąpiło w esencjonalność. Tu nie dzieje się nic na darmo, nic nie zdarza się bez głębokiego sensu i bez premedytacji. Tu zdarzenia nie są efemerycznym fantomem na powierzchni, tu mają one korzenie w głębi rzeczy i sięgają istoty. Tu rozstrzyga się coś każdej chwili, egzemplarycznie i po wszystkie czasy. Tu dzieją się wszystkie sprawy raz jeden tylko i nieodwołalnie“ (Schulz 1989, 326). Aus der Erzählung *Republika marzeń* (*Die Republik der Träume*).

Dem Kind ist selbstverständlich, daß, was es an seinem Lieblingsstädtchen entzückt, nur dort, ganz allein und nirgends sonst zu finden sei; es irrt, aber sein Irrtum stiftet das Modell der Erfahrung, eines Begriffs, welcher endlich der der Sache selbst wäre, nicht das Armselige von den Sachen abgezogene.<sup>16</sup> (Adorno 1973, 366)

### 3 Kindliche Erfahrung zwischen Nachglanz und Vorschein

Als nicht begriffliche Erkenntnis, die zu einer anderen Ordnung der Wirklichkeit vordringt, hat kindliche Erfahrung sowohl bei Schulz als auch bei Adorno Züge transzendenter bzw. „metaphysischer Erfahrung“ (Adorno 1973, 366). Die Sinndimension, die sich aus dem Potenzial kindlicher Erfahrung ergibt, deutet bei Schulz und Adorno jedoch in entgegengesetzte Richtungen. Für Schulz führt die kindliche Erfahrung zurück zu einem ursprünglichen Sinn der Dinge und damit zum Mythos. Für Adorno besteht der Wert kindlicher Erfahrung darin, dass in ihr die Verhältnisse einer künftigen veränderten Seinsordnung sichtbar werden, in ihr also ein Vorschein von Utopie liegt. Geradezu programmatisch stellen Schulz und Adorno diese beiden unterschiedlichen Perspektiven an zwei in ihrer Bildlichkeit erstaunlich ähnlichen Szenen dar.

In einer sommerlichen Phantasie versetzt sich der Ich-Erzähler in Schulz' Erzählung *Republika marzeń* (*Die Republik der Träume*) zurück in seine Kindheit. Mit den Eltern geht es aus der Stadt hinaus, um der sommerlichen Hitze zu entfliehen. Gemeinsam mit seinen Kameraden überlegt der kindliche Protagonist, wie es wäre, von hier aus einfach weiterzugehen, „in ein Land, das niemand und Gott gehörte, in strittiges und neutrales Grenzgebiet, wo sich die Ansprüche der Staaten verloren“, um dort, „von den Erwachsenen unabhängig“, eine „Republik der Jungen“ auszurufen. Hier sollte eine „neue Hierarchie der Maße und Werte“ eingeführt werden: „Es schien uns, als müßten wir nur die Barrieren und Grenzen der Konventionen beseitigen, die alten Lager, in welchen der Lauf der menschlichen Dinge festlag [...]“. Das von den Kindern erträumte unabhängige Staatsgebiet setzt die restriktive und praktische Lebensordnung der Erwachsenenwelt außer Kraft, indem es an ihrer Stelle Phantasie und Poesie zum „neue[n] Lebensprinzip“ ernennt (sämtliche Zitate in diesem Abschnitt Schulz 2000b, 338).<sup>17</sup>

<sup>16</sup> Siehe auch Adorno (2003b, 19): „Amorbach [ist mir; KT] das Urbild aller Städtchen geblieben, die anderen nichts als seine Imitation.“

<sup>17</sup> „W tych dniach dalekich powzięliśmy po raz pierwszy z kolegami ową myśl [...], ażeby powędrować jeszcze dalej, [...], w kraj już niczyj i boży, w pogranicze sporne i neutralne, gdzie gubiły się rubieże państw [...]. Tam chcieliśmy się oszańcować, uniezależnić od dorosłych, wyjść zupełnie poza obręb ich sfery, proklamować republikę młodych. Tu mieliśmy ukonstytuować prawodawstwo nowe i niezależne, wznieść nową hierarchię miar i wartości. Miało to być życie pod znakiem poezji i przygody [...]. Zdawało się nam, że trzeba tylko rozsunąć bariery i granice konwenansów, stare łożyska, w które ujęty był bieg spraw ludzkich,

Der zweckfreie, zur Phantasie hin geöffnete Umgang mit den Dingen im kindlichen Spiel hebt die engen Grenzen der nach praktischen Zwecken geordneten Wirklichkeit auf und stellt eine Verbindung zu jener Grundlage des Seins aller Dinge her, über die – so Schulz in einem Brief an den Schriftsteller Ignacy Witkiewicz – nicht hinauszukommen ist: zum Mythos.

Immer schon habe ich gefühlt, daß die Wurzeln des individuellen Geistes, genügend in die Tiefe gesenkt, sich in einer Art mythologischen Urwalds verlieren. Das ist der unterste Grund, über den man nicht mehr hinausgehen kann. [...] auf dem Grund aller menschlichen Ereignisse [bestehen; KT], wenn man sie aus den Hülsen der Zeit und Vielheit schält, bestimmte Urschemata [...], Geschichten, auf denen sich diese Ereignisse als große Wiederholungen formen.<sup>18</sup> (Schulz 2000a, 93)

Als Teil und als Überrest ursprünglicher Mythen, hat, laut Schulz, „jedes Fragment der Wirklichkeit [...] Anteil an einem universalen Sinn“ (Schulz 2000a, 240)<sup>19</sup>, in dem alle Prozesse des Seins zusammenlaufen. Das kindliche unbewusste Wissen um diese Teilhabe ist im späteren Leben für die meisten Menschen durch die alltägliche Lebenspraxis verstellt. Gleichwohl gibt es nach Schulz auch für den Erwachsenen die Möglichkeit eines Zugangs zum Mythos. Poesie, so machen die Lebensgrundsätze der „Republik der Jungen“ deutlich, ist die der kindlichen Erfahrung entsprechende Ausdrucksform, eine Form „wiedergekehrte[r] Kindheit“, wie Schulz das Ideal seiner Kunst im Brief an Pleśniewicz definierte. Als „metaphysisches Organ des Menschen“ (Schulz 2000a, 242)<sup>20</sup> versucht die Sprache einen Zugang zur sinnhaften Ordnung der Wirklichkeit wiederherzustellen. Wenn benennen bedeutet, Dinge einem universalen Sinn einzuschreiben (vgl. *ibid.*, 240), so müssen die Wörter jedoch erst wieder zu Trägern eigentlicher Bedeutung gemacht werden. Dies gelingt der Dichtung, indem sie, wie das phantasierende Spiel des Kindes, die Dinge aus ihren im Alltag festgelegten Bedeutungen heraushebt, sie für neue Verbindungen und Verwendungen öffnet und damit einen dynamisch universalen Zusammenhang aller Elemente der Wirklichkeit sichtbar macht. Das Wesen der

---

ażeby w życie nasze wlała się żywioł, wielki zalew nieprzewidzianego, powódź romantycznych przygód i fabuł“ (Schulz 1989, 329).

<sup>18</sup> „Zawsze czułem, że korzenie indywidualnego ducha, dostatecznie daleko w głąb ścigane, gubią się w mitycznym jakimś mateczniku. To jest dno ostateczne, poza które niepodobna już wyjść. [...] na dnie wszystkich zdarzeń ludzkich, gdy wyluskać je z plewy czasu i wielości, ukazują się pewne praszchematy, ‚historie‘, na których te zdarzenia formują się w wielkich powtórzeniach“ (Schulz 1989, 446). Schulz an Witkiewicz.

<sup>19</sup> „Każdy fragment rzeczywistości [...] ma udział w jakimś sensie uniwersalnym“ (Schulz 1989, 365). Dieses und das folgende Zitat aus Schulz' *Essay Mityzacja rzeczywistości* (1936; *Das Mythisieren der Wirklichkeit*).

<sup>20</sup> „Mowa jest metafizycznym organem człowieka“ (Schulz 1989, 367).

Wirklichkeit, ihre Identität, steckt nach Schulz in ihrer unablässigen Veränderung, die als Variation des Mythos zu begreifen ist. Die „Lockerung des Gewebes der Wirklichkeit“ (ibid., 92)<sup>21</sup> in der kindlichen Erfahrung und durch die dichterische Sprache dient somit gerade dem Rückgewinn der Sinneinheit der Wirklichkeit.

Die Rückkehr in die Erfahrungsweise des Kindes auf dem Weg über die poetische Sprache führt freilich letztlich in eine Aporie. Dies macht die Dichtung zu einer ambivalenten und komplexen Angelegenheit. Bietet sie einerseits ein Analogon zur kindlichen Erfahrung, so bleibt der Dichter doch insofern hinter ihr zurück, als das, was sich in der Kindheit intuitiv in aller „Fülle und Maßlosigkeit“ (ibid., 107)<sup>22</sup> erschloss, nur noch in „Scherben eines zerbrochenen Spiegels“ (Schulz 2000b, 118)<sup>23</sup> vor ihm liegt. Die Schulz' Werk durchziehende Ironie dient nicht zuletzt dazu, das Wissen des Dichters um die eigene Beschränktheit gegenüber der geforderten Aufgabe deutlich zu machen. Poetische Kreation ist selbst fragmentarisch, sie bleibt Re-Kreation. Die Spannung in Schulz' Werk liegt im Bewusstsein davon, dass einerseits die dichterische Schöpfung nur eine Kompensation für die „geniale Epoche“ der Kindheit sein kann, dass aber zugleich gerade der dichterischen Sprachschöpfung eine originäre sinn- und wirklichkeitsstiftende Funktion zukommt und somit die Sprache selbst möglicherweise der äußerste Grund ist, über den Sinnggebung nicht hinausgelangen kann, die Wirklichkeit also „Schatten des Worts“ ist (Schulz 2000a, 242).<sup>24</sup>

Malt sich Schulz' Erzähler in *Republika marzeń* einen unbesetzten Raum aus, so entdeckt Adorno als Kind einen solch scheinbaren Freiraum beim Herumstreifen zwischen den Dörfern des Odenwaldes und macht ihn sich im Spiel der Phantasie zu eigen:

Zwischen Ottorfszell und Ernsttal verlief die bayrische und badische Grenze. Sie war an der Landstraße durch Pfähle markiert, die stattliche Wappen trugen [...]. Reichlicher Zwischenraum zwischen beiden. Darin hielt ich mit Vorliebe mich auf, unter dem Vorwand, [...] jener Raum [...] sei frei, und ich könnte dort nach Belieben die eigene Herrschaft errichten. [...] Das Land aber, das sie [die Wappenpfähle; KT] umschlossen, und das ich, spielend mit mir selbst, okkupierte, war Niemandland. [...] Es [das Wort Niemandland; KT] ist aber die getreue Übersetzung des griechischen –

<sup>21</sup> „rozluźnieni[e] tkanki rzeczywistości“ (Schulz 1989, 445). Schulz 1935 an Witkiewicz.

<sup>22</sup> „powtórnie dzieciństwo, jeszcze raz mieć jego pełnię i bezmiar“ (Schulz 1989, 424). Schulz 1936 im Brief an Pleśniewicz.

<sup>23</sup> „ułamki potłuczonego zwierziadła“ (Schulz 1989, 119). Aus der Erzählung *Księga* (*Das Buch*).

<sup>24</sup> „rzeczywistość jest cieniem słowa“ (Schulz 1989, 368). Aus *Mityzacja rzeczywistości*. Schulz pflegt insgesamt einen ironisch-spielerischen Umgang mit dem Mythos, den er individualisiert und privatisiert. Die Gültigkeit mythischen Denkens hebt dies jedoch nicht auf. Zu Schulz' Umgang mit dem Mythos siehe z. B. Schulz an Witkiewicz (Schulz 2000b, 93), Prokopczyk (1999).

Aristophanischen –, das ich damals besser verstand, je weniger ich es konnte, Utopie. (Adorno 2003b, 19–20)

In einem Brief an Thomas Mann definiert Adorno 1945 „Utopie“ als den „Traum einer von Zwecken nicht entstellten Welt“ (Adorno & Mann 2002, 16).<sup>25</sup> Die vom Kind im Spiel geschaffenen Räume, in denen die Zweck-Mittel-Rationalität als herrschendes Prinzip ausgesetzt ist, erlauben den Blick auf eine Wirklichkeit, in der ein grundsätzlich verändertes Verhältnis ihrer Elemente zueinander besteht. Ist einerseits für Erwachsene in der Erinnerung an die Erfahrungsweise des Kindes ein Vorgeschmack dieser anderen Wirklichkeit bewahrt, so ist andererseits kindliche Erfahrung nicht selbst schon die Verwirklichung der Utopie, die Adorno als Zielzustand einer von gesellschaftlicher Entfremdung befreiten Welt ansetzt. Denn, so Adorno, die Natur des kindlichen Spiels liegt gerade darin, dass es nicht real ist: „Die Unwirklichkeit der Spiele gibt kund, daß das Wirkliche es noch nicht ist. Sie sind bewußtlose Übungen zum richtigen Leben“ (Adorno 2003a, 260–261).

Kindliche Erfahrung kann nur ein Vorschein sein, sie ist ein Versprechen auf eine andere Wirklichkeitsordnung.<sup>26</sup> Inhaltlich jedoch muss die Utopie nach Adorno unbestimmt bleiben. Sie „auszupinseln“ würde bedeuten „das ganz ‚Andere‘ mit dem Maß des Bestehenden zu messen“ (Schweppenhäuser 1996, 93). Die Erkenntnis einer künftigen veränderten Wirklichkeit ist für Adorno nur negativ, in der Kritik am Bestehenden zu haben. Erfahrungen und Lebensmomente, in denen die verfestigten Strukturen des Bestehenden aufgebrochen werden – wie beispielsweise im kindlichen Spiel – ermöglichen eine solche Kritik.<sup>27</sup> Während für Schulz die in der Dichtung wenngleich fragil und fragmentarisch wiedereingeholte Kindheit eine Kompensation für die bestehende Wirklichkeit darstellt, hat Kindheitserinnerung als Versprechen auf ein Künftiges für Adorno keine kompensatorische Funktion. Vom gesellschaftskritischen Standpunkt her wäre solche Kompensation denn auch kontraproduktiv, würde sie als Fluchtraum eine tatsächliche Veränderung doch eher behindern. Erkenntnistheoretisch betrachtet sensibilisiert die Kindheitserinnerung die Menschen jedoch für eine veränderte Wahrnehmung der Wirklichkeit.

<sup>25</sup> Brief an Thomas Mann vom 03.06.1945. Adorno nennt Mann in seinem Brief als beispielhaft für sein Verständnis von Utopie. Interessanterweise bezieht sich Schulz gleichfalls auf Mann, um in einem Brief gegenüber Witkiewicz seine Gedanken der mythologischen Grundlage der Wirklichkeit zu veranschaulichen (vgl. Schulz 2000c, 93). Wie Adorno, so pflegte auch Schulz mit Thomas Mann einen Briefwechsel, der, Dutsch nach, aber verloren ging (Schulz 2000a, 361).

<sup>26</sup> Zur kindlichen Erfahrung als Versprechen bei Adorno, siehe Thaidigsmann (2009, 123–128).

<sup>27</sup> Durch ein Merkmal qualifiziert Adorno die Utopie dann aber doch genauer. Sie sei, so der Philosoph, durch „Hingabe“ bestimmt – ein Merkmal, das ganz kindlicher Wirklichkeitserfahrung entspricht (vgl. Habermas 1981, 165).

Wenn für Adorno auch nicht wie für Schulz die Kindheit in der Kunst wieder eingeholt werden kann, so sind doch auch für ihn Kunst und Kindheitserfahrung einander verwandt. Gleich der kindlichen Wirklichkeitserfahrung bietet nach Adorno auch authentische Kunst eine eigene Erkenntnisform, die, da sie nicht begrifflich fixiert, die Wirklichkeit aus der durch praktische Zwecke „entstellten“ Wahrnehmung herauslöst. Wie die Erinnerung an die Kindheit ist das Kunstwerk ein Splitter in der bestehenden Realität, der einen utopischen Vorschein auf eine veränderte Wirklichkeit gibt.<sup>28</sup> Die Herausforderung für die kritische Philosophie besteht laut Adorno nun darin, in begrifflicher Sprache unverkürzte, individuelle Erfahrung, wie sie in Erinnerung an die Kindheit und Kunst aufscheint, nicht preiszugeben, sondern ihr gerecht zu werden (vgl. Schweppenhäuser 1996, 66).<sup>29</sup> Insbesondere die Aphorismensammlung *Minima Moralia*, doch auch Adornos Prosaskizzen über seine Kindheit in Amorbach sind Versuche einer Erkenntnis aus der Verbindung von individueller Erfahrung und Begrifflichem. Als solche sind die Kindheitsskizzen, nicht, wie Adorno von der Kritik mitunter vorgeworfen wurde, belanglose „feuilletonistische Plauderei“, sondern, so Adorno selbst, „sehr gewagt“ (Adorno 2003b, 215).<sup>30</sup>

In dem bereits oben zitierten Brief an Witkiewicz fragt sich Schulz, wie wohl eine philosophische Interpretation seiner Erzählungen *Sklepy cynamonowe* aussehen könnte und stellt dabei die Eigenheiten künstlerischer und philosophischer Erkenntnis einander gegenüber. Im Kunstwerk, so Schulz, sei „noch nicht die Nabelschnur gelöst, die es mit der Gesamtheit unserer Problematik verbindet“ (Schulz 2000a, 91), während die philosophische Interpretation mit ihren rationalisierenden, analytischen Methoden „nur mehr Ausschnitte aus der Gesamtheit der Problematik“ heraushebe und erfasse. „[D]as philosophische Credo der *Zimtläden*“, so Schulz, wäre daher „eher der Versuch einer Be-

<sup>28</sup> Zur Erkenntnisfunktion der Kunst bei Adorno siehe seine ästhetische Theorie (v.a. in Adorno: *Ästhetische Theorie*, 1970). Siehe auch Schweppenhäuser (1996, 117–119).

<sup>29</sup> Adorno bemüht sich in seiner Philosophie darum, die Reduktion des Sinns, die durch das Denken in Begriffen entsteht, durch ein „Denken in Konstellationen“ aufzubrechen (vgl. Schweppenhäuser 1996, 66–67). Möglicherweise versucht er damit – wenngleich in anderer Perspektive – auf begrifflicher Ebene zu realisieren, was Schulz als das Charakteristikum nicht-begrifflicher, dichterischer Sprache versteht, nämlich „das Leistungsvermögen der Worte durch Kurzschlüsse wieder her[zustellen], die aus Kumulation entstehen“ (Schulz 2000a, 242) / „Poeta przywraca słowom przewodnictwo przez nowe spięcia, które z kumulacji powstają“ (Schulz 1989, 368).

<sup>30</sup> Dort auch zur Kritik an Adornos Amorbach-Skizzen. Gewagt sind die Amorbach-Skizzen für Adorno auch deshalb, weil er sich in ihnen eine Nähe zum „Idyllischen“ (ibid.) zugesteht und die Welt seiner Kindheit immer wieder in ein geradezu mythologisches Licht taucht, was durchaus tröstende Bedeutung hat (ibid., 22). In seinen kritischen Schriften, insbesondere der mit Horkheimer herausgegebenen *Dialektik der Aufklärung* schreibt er dem Mythos hingegen eine negativ die Unveränderlichkeit der Welt zementierende Funktion zu. Zur Nähe von Adornos Kindheitsbeschreibungen und Mythos vgl. auch Jäger (2009, 13), der darauf hinweist, dass Adorno es liebte, seine eigene Herkunft aus einer mythologischen Ahnenreihe herzuleiten. Dies erinnert an Schulz' Erklärung gegenüber Witkiewicz, er habe sich einen privaten mythologischen Stammbaum geschaffen (Schulz 2000a, 93).



schreibung der dort gegebenen Wirklichkeit als deren Begründung [...]“ (ibid.). Die Erzählungen von *Sklepy cynamonowe* vermitteln jedoch nicht das Bild einer bestimmten Wirklichkeit, sondern umreißen deren Struktur: Sie „geben ein bestimmtes Rezept für die Wirklichkeit“ (ibid.).<sup>31</sup> Die Bedeutung von Erkenntnis, wie sie in der Erfahrung des Kindes und im Kunstwerk zugänglich wird, liegt sowohl für Schulz als auch für Adorno darin, Schlüssel für eine Wahrnehmung von Wirklichkeit zu sein, die in Schulz’ mythologischer Perspektive im alltäglichen Getriebe der Welt nicht mehr und in Adornos utopischer Perspektive noch nicht realisiert ist.

#### 4 Schluss – Die Welt aus der Perspektive der Erlösung

Kindliche Erfahrung bietet für Schulz wie auch für Adorno die Möglichkeit eines Blicks auf die Welt aus der Perspektive ihrer Erlösung. Beide binden den Gedanken der Erlösung an den jüdisch konnotierten Begriff des Messianischen, den sie, wie bereits beim Umgang mit den Feiertagen, in seiner Grundbedeutung übernehmen, während sie ihn zugleich aus seiner religiösen Verhaftung herauslösen.<sup>32</sup> Das Erreichen eines Zustandes der Erlösung bleibt jedoch bei beiden ungewiss: bei Schulz, weil die Erlösung, wie das Bild der zurückgeholten Kindheit verdeutlicht, an sich unmöglich ist, bei Adorno, weil das utopische „Noch nicht“ nur im negativen Vorschein der Kritik zugänglich ist. Die Hoffnung auf Realisierung der Erlösung steht sowohl für Schulz als auch für Adorno hinter der Erkenntnis von der Erlösungsbedürftigkeit der Wirklichkeit, die die mythologische respektive utopische Perspektive kindlicher Erfahrung bietet, zurück.<sup>33</sup> Diese Erkenntnis wird selbst produktiv: für Schulz in schöpferisch-poetischer Kreativität, für Adorno in kritischer Gesellschaftsanalyse. Was bei Adorno „in den Ritzen und Brüchen dessen [...], was ist“ als utopisches Versprechen einer veränderten Wirklichkeit nistet (Thaidigsmann 2009, 134), wächst in Schulz’ Prosa ganz bildlich als mythologischer Urgrund aus solchen Ritzen hervor.

<sup>31</sup> „Sądzę, że zracjonalizowanie widzenia rzeczy tkwiącego w dziele sztuki jest [...] zubożeniem problematyki dzieła. [...]. W dziele sztuki nie została jeszcze przerwana pępowina łącząca je z całością naszej problematyki [...]. W filozoficznej interpretacji mamy już tylko wypruty z całości problematyki preparat anatomiczny. Mimo to sam ciekaw jestem, jak brzmiałoby w formie dyskursywnej credo filozoficzne *Sklepów cynamonowych*. Będzie to raczej próba opisu rzeczywistości tam danej aniżeli jej uzasadnienie. *Sklepy cynamonowe* dają pewną receptę na rzeczywistość [...]“ (Schulz 1989, 444).

<sup>32</sup> Siehe in *Minima Moralia* das Prosastück *Zum Ende* (Adorno 2003a, 283) und Schulz’ Brief an Pleśniewicz (Schulz 2000a, 107). Der Gedanke des Messianischen verbindet Schulz’ Werk auch in besonderer Weise mit dem Denken Walter Benjamins, eine Parallele, die in der Schulz-Forschung bereits mehrfach thematisiert wurde, vgl. den Forschungsüberblick in Czabanowska-Wróbel (2016).

<sup>33</sup> Vgl. zu dieser Frage Adorno (2003a, 283) und Schulz’ Erzählung *Księga* (Schulz 2000b, 117).

Adorno und Schulz sind beide Kinder der Moderne. Während für Adorno die Erfahrung der Umbrüche der Moderne mit ihren Prozessen der Rationalisierung und Kommerzialisierung zu einem Drehpunkt seines Denkens wird, kann man diese Umbrüche bei Schulz zumindest als einen Verstärker für die Entwicklung seiner mythologisierend-poetischen Wirklichkeitskonzeption verstehen. Wahrnehmungskritik und Erkenntnis stehen bei Adorno wie auch bei Schulz in engem Zusammenhang mit Sprache und Sprachkritik. Obgleich Schulz' Prosa nicht als Gesellschaftskritik konzipiert ist und auch nicht als solche verstanden werden will, sensibilisiert sie für den reduktionistischen Umgang mit der Wirklichkeit, wie er gerade in der praktischen Zurichtung der Sprache zum Ausdruck kommt. Die moralischen Implikationen von Adornos Perspektive auf kindliche Erfahrung als nicht-reduktionistische Erkenntnis haben in Schulz' poetischer Konzeption ihr ästhetisches Pendant im Aufbrechen der „Langeweile“. Das funktionalistisch Gleichgemachte ist langweilig und damit ästhetisch uninteressant. In ethischer Perspektive wird diese Langeweile zugleich als Symptom eines Umgangs mit den Dingen durchsichtig, der diese ihres individuellen Werts beraubt. Schulz von Adorno her zu lesen lenkt die Aufmerksamkeit auf die ethischen Implikationen im Werk des polnisch-jüdischen Schriftstellers.

Durch die Totalitarismen des 20. Jahrhunderts haben mythologisches wie utopisches Denken eine tiefe Diskreditierung erfahren. Beiden begegnet Adorno daher mit großer Vorsicht. Die Differenz zwischen Mythologie bei Schulz und kritisch gewendeter Utopie bei Adorno spiegelt sich auch in den Formen ihres Schreibens. Bei Schulz verbinden sich die „Scherben“ des Mythos zu einem poetischen Universum, die Splitter nicht-identischer, auf Utopisches vorausdeutender Erfahrung bei Adorno bleiben hingegen als einzelne Aphorismen und Skizzen unverbunden. Für Schulz bedeutet das Ideal ganz „zur Kindheit ‚heranzureifen‘“ „die wirkliche Reife“ (Schulz 2000a, 107).<sup>34</sup> Adorno habe, so Jürgen Habermas „die Alternative von Kindbleiben oder Erwachsenwerden nie akzeptiert“ (Habermas 1981, 170). Die Verbindung von Kindlichem und Erwachsenem steht bei ihm für die Verbindung von Aufgeschlossenheit für Erfahrung und Festhalten am Erkenntnispotenzial der Vernunft (vgl. dazu Kirchhoff 2004, 83). Adorno und Schulz sind unbestritten zwei der großen „reifen Unreifen“ des zwanzigsten Jahrhunderts, die gerade aus kindlich reifer Unreife heraus – jeder auf seine eigene Art – besondere Zugänge zur Erfahrung der Wirklichkeit eröffnen.

---

<sup>34</sup> „Moim ideałem jest ‚dojrzeć‘ do dzieciństwa. To by dopiero była prawdziwa dojrzałość“ (Schulz 1989, 424). Schulz im Brief an Pleśniewicz.

## Literatur

- Adorno, Theodor W. 1973. *Negative Dialektik. Jargon der Eigentlichkeit*. Hrsg. v. R. Tiedemann. *Gesammelte Schriften*, Bd. 6. Frankfurt a. M.: Suhrkamp.
- Adorno, Theodor 1986. *Vermischte Schriften I*. Hrsg. v. R. Tiedemann. *Gesammelte Schriften*, Bd. 20.1. Frankfurt a. M.: Suhrkamp.
- Adorno, Theodor 2003a. *Minima Moralia. Reflexionen aus dem beschädigten Leben*. Hrsg. v. R. Tiedemann. *Gesammelte Schriften*, Bd. 4. Frankfurt a. M.: Suhrkamp.
- Adorno, Theodor 2003b. *Kindheit in Amorbach. Bilder und Erinnerungen*. Hrsg. v. R. Pabst. Frankfurt a. M./Leipzig: Insel.
- Adorno, Theodor W. und Thomas Mann. 2002. *Briefwechsel 1943–1955*. Hrsg. v. C. Gödde und T. Sprecher. *Theodor W. Adorno: Briefe und Briefwechsel*, Bd. 3. Frankfurt a. M.: Suhrkamp.
- Bolecki, Włodzimierz, Jerzy Jarzębski und Stanisław Rosiek. 2003. *Słownik schulzowski*. Danzig: Slowo/obraz/terytoria.
- Budurowycz, Bohdan. 1986. „Galicia in the Work of Bruno Schulz.“ *Canadian Slavonic Papers*, Bd. 28: 359–368.
- Czabanowska-Wróbel, Anna . 2016. „Drohobyckie i berlińskie dzieciństwo. Bruno Schulz i Walter Benjamin.“ *Ruch Literacki* 57 (1): 45–58.
- Ficowski, Jerzy. 2003. *Regions of the Great Heresy. A Biographical Portrait*. New York/London: W. W. Norton & Company.
- Habermas, Jürgen. 1981. „Theodor W. Adorno“. In Ders. *Philosophisch-politische Profile*. Frankfurt a. M.: Suhrkamp: 160–179.
- Jarzębski, Jerzy. 2003. „Schulz i dramat tworzenia.“ *Teksty Drugie*. Nr. 5: 9–16.
- Jäger, Lorenz. 2009. *Adorno. Eine politische Biographie*. München: Pantheon.
- Kirchhoff, Christine. 2004. „Die Möglichkeit als eine der Wirklichkeit fassen. Über den Erfahrungsbegriff Theodor W. Adornos.“ In *Gesellschaft als Verkehrung. Perspektiven einer neuen Marx-Lektüre. Festschrift für Helmut Reichelt*, hrsg. v. C. Kirchhoff u.a., 83–103. Freiburg: ça ira.
- Koch, Klaus u.a. (Hrsg.). 2000. *Reclams Bibellexikon*. 6. verb. Aufl. Stuttgart: Reclam.
- Philipsen, Peter-Ulrich. 2007. „Kinderbild der Moderne. Adornos Der Schatz des Indianer-Joe“. In *Geschichte – Kultur – Bildung. Philosophische Denkrichtungen. Johannes Rohbeck zum 60. Geburtstag*, hrsg. v. P. Breitenstein, und V. Steenblock und J. Siebert, 103–116. Hannover: Siebert.
- Posnock, Ross. 2006. *Philip Roth's Rude Truth: the Art of Immaturity*. Princeton: Princeton University Press.
- Prokopczyk, Czesław Z. 1999. „The Mythical and the Ordinary in Bruno Schulz.“ In *Bruno Schulz: New Documents and Interpretations*, hrsg. v. Cz. Prokopczyk, 175–209. New York u. a.: Peter Lang.
- Schulz, Bruno. 1989. *Opowiadania. Wybór esejów i listów*. Hrsg. v. J. Jarzębski. Wrocław: Ossolineum.
- Schulz, Bruno. 2000a. *Die Wirklichkeit ist Schatten des Wortes. Aufsätze und Briefe*. Hrsg. v. J. Ficowski. Übersetzt von Mikolaj Dutsch und Joseph Hahn. München: dtv.

- Schulz, Bruno. 2000b. *Die Zimtläden und alle anderen Erzählungen*. Hrsg. v. M. Dutsch. Übersetzt von Joseph Hahn. München: dtv.
- Schweppenhäuser, Gerhard. 1996. *Theodor W. Adorno zur Einführung*. Hamburg: Junius.
- Thaidigsmann, Edgar. 2009. „Das Versprechen. Metaphysische Erfahrung bei Theodor W. Adorno.“ *Zeitschrift für Theologie und Kirche*. 106. Heft 1. März: 118–136.
- Tiedemann, Rolf. 1993. „Begriff, Bild, Name. Über Adornos Utopie der Erkenntnis.“ In *Frankfurter Adorno-Blätter II*, 92–111. München: edition text + kritik.