

Die Bewertung von Kohärenz und Kohäsion in narrativen DaF-Texten: Eine korpusbasierte Untersuchung sprachlicher Einflussfaktoren

Helena Wedig & Carola Strobl

Abstract: Die Bewertung von schriftlichen Leistungen im Rahmen des DaF-Unterrichts wird oft auf Basis von numerischen Skalen mit vorgegebenem Bewertungshorizont durchgeführt. Trotz der weiten Verbreitung dieser Bewertungsform sind die textuellen Merkmale, die ExpertInnen bei der Bewertung beeinflussen, noch relativ unbekannt. Eine skalenbasierte Bewertungsform lag auch der in diesem Beitrag vorgestellten Analyse eines longitudinalen DaF-Korpus von 60 narrativen Texten mit einem Bewertungsfokus auf Kohärenz und Kohäsion zugrunde. Um zu eruieren, inwieweit sich die manuellen Bewertungen mithilfe von über 200 automatisch ermittelten sprachlichen Merkmalen vorhersagen lassen, wurde eine Korrelationsanalyse durchgeführt und ein lineares Regressionsmodell ermittelt. Es zeigt sich, dass als kohärent bewertete Texte u.a. einen hohen Grad semantischer und konjunkionaler Verknüpfungen sowie eine erhöhte Koreferenz in Form von Stammüberlappungen aufweisen. Allerdings gab es Texte, die von diesem Modell abweichen. Um mögliche Ursachen dafür zu finden, wurden einige Texte einer vertiefenden Analyse hinsichtlich koreferenziellen Relationen auf der Diskursebene unterzogen, was sich als vielversprechender Erklärungsfaktor erwies. Die Ergebnisse dieser Studie werden mit Vorgängerstudien zum Englischen als Fremdsprache verglichen und mit Bezug auf die Unterrichtspraxis diskutiert.

The assessment of written performance in German as a foreign language (GFL) instruction is often based on numerical scales according to descriptive rubrics. Despite the widespread use of human rubrics-based assessment, little is known about the linguistic features that have an impact on human rating. The present study analyses a longitudinal GFL corpus consisting of 60 texts that were rated along rubrics for cohesion and coherence. The human ratings were correlated with over 200 automatically computed linguistic features by means of a correlation analysis and a linear regression model. The results show that texts with a high cohesion score have a high degree of semantic and conjunctive relatedness and an increased use of coreferential cohesion through overlapping stems. A number of texts, however, did not fit this model. To investigate additional explanatory factors, an in-depth analysis of several texts was conducted, targeting the coreference relations at text level. This approach yielded promising results regarding its explanatory power. The results are compared with previous research on English as a FL and implications for writing pedagogy are discussed.

Schlagwörter: Kohäsion; Kohärenz; sprachliche Merkmale; Korrelation; DaF; Deutsch; human rating; cohesion; coherence; German as a foreign language.

Wedig, Helena & Strobl, Carola (2022):
Die Bewertung von Kohärenz und Kohäsion in narrativen DaF-Texten:
Eine korpusbasierte Untersuchung sprachlicher Einflussfaktoren.
Zeitschrift für Interkulturellen Fremdsprachenunterricht 27: 1, 369–396.
<http://tujournals.ulb.tu-darmstadt.de/index.php/zif>

1 Einführung

Mithilfe moderner sprachtechnologischer Verfahren können Zusammenhänge zwischen automatisch erhobenen sprachlichen Merkmalen und manuellen Bewertungen von Texten hergestellt werden, um festzustellen, von welchen Merkmalen sich die BewerterInnen leiten lassen. Eine eingehende Analyse dieser Einflüsse erlaubt nicht nur die Optimierung des Fremdsprachenunterrichts z.B. durch Fokus auf kohäsive Mittel, welche die durch BewerterInnen wahrgenommene Kohärenz und Kohäsion eines Textes deutlich erhöhen, sondern auch Rückschlüsse auf die Performanz und Reliabilität automatischer Bewertungssysteme von Lernertexten (vgl. Crossley 2020: 416; Guo/Crossley/McNamara 2013: 221–222; Jung/Crossley/McNamara 2019: 38; Crossley/Kyle/Allen/Guo/McNamara 2014: 8). So zeigt sich u.a., dass es BewerterInnen nur selten möglich ist, ihre Fokuspunkte der Analyse bzw. die Gründe der Bewertung detailliert zu benennen (vgl. Lumley 2005: 311).

Besonders geeignet für eine Analyse des Zusammenhangs zwischen automatisch erhobenen Merkmalen und manuellen Bewertungen ist das Bewertungskriterium der Kohärenz und Kohäsion, da diese übergreifenden Textmerkmale weder auf einen bestimmten Inhalt noch auf die sprachliche Korrektheit der Textproduktion Bezug nehmen. De Beaugrande und Dressler (1981: 3–5) definieren Kohäsion als Textkriterium, welches die Art betrifft, „wie die Komponenten des Oberflächentextes, d.h. die Wörter, wie wir sie tatsächlich hören oder sehen, miteinander verbunden sind“ (de Beaugrande/Dressler 1981: 3–4). Sie stellen fest, dass „Kohäsion auf grammatischen Abhängigkeiten beruht“ (de Beaugrande/Dressler 1981: 4). Kohärenz bezeichnen sie als die Funktion betreffend, „durch die die Komponenten der Textwelt, d.h. die Konstellation von Konzepten (Begriffen) und Relationen (Beziehungen), welche dem Oberflächentext zugrunde liegen, füreinander gegenseitig zugänglich und relevant sind“ (de Beaugrande/Dressler 1981: 5). Obwohl beide Aspekte unterschiedliche Fokuspunkte der Textanalyse darstellen, werden sie in der Fremdsprachenlehre und bei der Bewertung von Texten oft zusammengefasst, so z.B. auch in dem in dieser Studie verwendeten Bewertungsraster nach Kuiken und Vedder (2017).

Bisher gibt es keine Studie, die die Korrelation zwischen sprachlichen Merkmalen und der manuellen Bewertung von Kohärenz und Kohäsion explizit bezogen auf den DaF-Kontext analysiert, wohingegen für das Englische als Lernaltersprache bereits einige Forschungsergebnisse vorliegen (z.B. Crossley/McNamara 2011). Dieser Beitrag stellt einen Ansatz dar, diese Forschungslücke zu schließen. Dies geschieht anhand einer Analyse eines longitudinalen DaF-Korpus auf der Basis von über 200 automatisch erhobenen sprachlichen Merkmalen und deren Korrelation zu der Bewertung dreier erfahrener BewerterInnen. Diese Bewertung bildet den

Hauptfokus der vorliegenden Studie. Einige Texte, die nicht den allgemeinen Tendenzen bei den Resultaten dieser quantitativen Analyse entsprechen, wurden einer zusätzlichen manuellen Analyse bezüglich Koreferenzstrukturen auf Diskursebene unterzogen, um mögliche Gründe für die Abweichungen zu eruieren.

Dieser Beitrag ist wie folgt aufgebaut: Zunächst wird im theoretischen Rahmen ein Forschungsüberblick über den Stand der automatischen Erfassung von sprachlichen Merkmalen der Kohärenz und Kohäsion gegeben und in Zusammenhang mit der manuellen Bewertung auf der Basis von deskriptiven Skalen gesetzt. Im dritten Kapitel werden die Datengrundlage, der Bewertungshorizont und das methodische Vorgehen beschrieben. Die Ergebnisse der Analysen werden im vierten Kapitel dargestellt. Die letzten beiden Kapitel bieten eine Einordnung der Ergebnisse in den Forschungszusammenhang sowie ein Fazit mit Ausblick auf mögliche Konsequenzen für den Fremdsprachenunterricht sowie weitere Forschungsdesiderata.

2 Methoden der Qualitätsanalyse fremdsprachlichen Schreibens

Die Bewertung der Schreibkompetenz von Zweit-/ und Fremdsprachen-LernerInnen (L2-LernerInnen) kann sowohl quantitativ als auch qualitativ vorgenommen werden. Dabei müssen verschiedene Aspekte der Textqualität berücksichtigt werden. Ein viel genutztes Modell für die Analyse von Lerneräußerungen ist das sogenannte CAF-Modell nach Wolfe-Quintero, Inagaki und Kim (1998), das nach den Aspekten Komplexität (*complexity*), Korrektheit (*accuracy*) und Flüssigkeit (*fluency*) unterscheidet. Die Kategorie der Korrektheit bezieht sich auf die Abweichung von der Norm, hier werden also alle Arten von Fehlern, von der Rechtschreibung über Grammatik hin zur Semantik, analysiert. Flüssigkeit wird im Hinblick auf die Textproduktion in zeitlich eingeschränktem Rahmen als Dichte produzierter linguistischer Einheiten (auch: *speed fluency*) definiert. Komplexität umfasst die lexikalische und syntaktische Komplexität der Sprachproduktionen sowohl im Hinblick auf die Ausgereiftheit von Wortschatz und syntaktischen Strukturen als auch im Hinblick auf die Variation der verwendeten Begriffe und Strukturen (vgl. Housen/Kuiken 2009: 3–5). Das CAF-Modell diente u.a. als Basis für die Erstellung automatischer Qualitätsanalysen von Lernertexten. Im CAF-Modell sind jedoch nicht alle textuellen Qualitätsmerkmale erschöpfend erfasst. Dies führte zu einer Erweiterung des Modells durch die Hinzufügung von funktionalen Aspekten auf der Basis der Systemisch Funktionalen Linguistik (*functionality*) (CAFF) (vgl. Link 2015: 35). Eine andere Definition von funktionalen Aspekten liegt bei der manuellen Bewertungsskala für Lerneräußerungen aufgrund von *funktionaler Adäquatheit* (*functional adequacy*, FA) nach Kuiken und Vedder (2017) vor, wo die

sprachliche und inhaltliche Erfüllung der Erwartungshaltung hinsichtlich der Aufgabenstellung und des Textgenres im Vordergrund steht.

Im Folgenden wird ein kurzer Überblick über bestehende automatische sowie manuelle Ansätze geboten, um Kohärenz und/oder Kohäsion in Lernertexten zu bewerten.

2.1 Automatische Analyse von Textqualität mit Fokus auf Kohäsion und Komplexität

Das Textverständnis und der Lesefluss werden neben anderen kontextgebundenen Faktoren auch durch mehrere textinterne Faktoren beeinflusst, z.B. komplexe Satzstrukturen, erhöhte lexikalische Diversität sowie Verwendung seltener Wörter. Diese quantifizierbaren Eigenschaften der Komplexität beeinflussen gemeinsam mit den Merkmalen der Textkohäsion wie z.B. der Verwendung von Konnektiven, Referenzen sowie semantisch verwandten Begriffen die Verständlichkeit bzw. Lesbarkeit eines Textes (vgl. McNamara/Crossley/McCarthy 2010: 63). Daher werden in Tools zur automatischen Bewertung von Kohäsion auch eine Vielzahl von Komplexitätsmerkmalen integriert. Der technische Fortschritt im Rahmen der Korpus- und Computerlinguistik ermöglicht die Betrachtung einer vielfältigen Menge an automatisch erhobenen Eigenschaften. Für das Englische existieren bereits eine Vielzahl an Tools, die es ermöglichen, Texte automatisch hinsichtlich bis zu 400 sprachlicher Eigenschaften der Komplexität und Kohäsion zu analysieren (zur Übersicht: Vyatkina/Housen 2020: 324–326). Im Folgenden wird eine Auswahl von Tools für das Englische sowie ein Tool für das Deutsche kurz vorgestellt:

TAALES, TAASSC, TAALED, TAACO & Coh-Matrix

Das Tool für die automatische Analyse lexikalischer Gewandtheit (*sophistication*) (TAALES: Kyle/Crossley 2015; Kyle/Crossley/Berger 2018) erlaubt die automatische Analyse von 420 Texteigenschaften. Es beinhaltet u.a. die Wortfrequenz, die spezifische Analyse der in akademischer Sprache vorkommenden Ausdrücke, eine Analyse psycholinguistischer Wortinformationen sowie kontextuelle Besonderheiten. Die syntaktische Gewandtheit und Komplexität (z.B. in Form von Phrasenstruktur) kann mithilfe des Tools TAASSC (Kyle 2016) analysiert werden. TAALED erlaubt die Analyse der lexikalischen Diversität (Kyle/Crossley/Jarvis 2021) und TAACO die spezifische Analyse der vorhandenen kohäsiven Mittel (Crossley/Kyle/Dascalu 2019; Crossley/Kyle/McNamara 2016). Letzteres umfasst dabei sowohl lokale als auch globale Eigenschaften der Kohäsion, wie z.B. Argumentüberlappungen, *Latent Semantic Analysis* (LSA) und Synonyme. Das Tool Coh-Matrix (Graesser/McNamara/Louwerse/Cai 2004; McNamara/Louwerse/McCarthy/Graesser 2010) stellt umfangreiche Indizes zur Analyse der Textkohärenz und -kohäsion zur Verfügung. So ist es möglich, englische Daten u.a. mittels deskriptiver

Statistik, Worteigenschaften, lexikalischer Diversität, syntaktischer Analysen, ko-referenzieller Kohäsion, LSA, Konnektive sowie Lesbarkeitsindizes zu analysieren.

CTAP

Für das Deutsche gibt es bisher nur ein Tool, welches die kollektive Analyse einer Vielzahl von Komplexitäts- und Kohäsionseigenschaften erlaubt. Die *Common Text Analysis Plattform* (CTAP) (Chen/Meurers 2016) verfügt über die Möglichkeit, zwischen über 300 Eigenschaften zu wählen. Dazu gehören z.B. Eigenschaften der lexikalischen, syntaktischen und morphologischen Komplexität, des Diskurses sowie der semantischen Kohärenz der Texte (Hancke/Vajjala/Meurers 2012: 1066–1072; Hancke 2013: 27–46; Weiß 2017: 68–104). Es besteht dabei eine große Schnittmenge zu den durch Coh-Metrix (Graesser et al. 2004; McNamara et al. 2010) und TAACO (Crossley/Kyle/Dascalu 2019; Crossley/Kyle/McNamara 2016) zur Verfügung gestellten Eigenschaften, die auf das Deutsche abgebildet wurden. Viele der durch das Tool zur Verfügung gestellten Merkmale der Komplexität (z.B. Art der Konnektive, lexikalische Varianz sowie Stammüberlappungen) sind auch bei der Analyse der Textkohäsion von Bedeutung.

2.2 Manuelle Bewertungen von schriftlichen Lerneräußerungen

Um einen Einblick in die schriftsprachliche Kompetenz von LernerInnen zu erlangen, werden im Rahmen des Fremdsprachenunterrichts Verschriftlichungen verlangt, die anschließend durch SprachexpertInnen bewertet werden. Diese Bewertungen werden durch eine vielfältige Anzahl an Einflüssen geprägt, die über die reine Textoberfläche hinausgehen. So werden sie durch Faktoren wie die zur Verfügung stehenden Bewertungsskalen und die Variabilität zwischen den BewerterInnen beeinflusst. Die Variabilität entsteht z.B. durch die Gewichtung einzelner Faktoren, die Erfahrung der BewerterInnen sowie ihr Training (vgl. Barkaoui 2010: 66–68; Kuiken/Vedder 2014: 337–341). Barkaoui (2010:68) zufolge nehmen die Bewertungsskalen den stärksten Einfluss auf die Bewertung der Verschriftlichungen. Er unterscheidet dabei zwischen analytischen (*analytic*) und ganzheitlichen (*holistic*) Verfahren. Analytische Verfahren beruhen auf untergeordneten Bewertungen individueller Eigenschaften, welche summiert die Gesamtbewertung darstellen und im Bewertungsprozess getrennt voneinander betrachtet werden müssen. Ganzheitliche Bewertungen beruhen auf einem festgelegten Wert für den gesamten Text.

Analytische Skalen bieten den Vorteil, dass sich BewerterInnen explizit auf die Qualitätsmerkmale der einzelnen Skalen stützen können, um die Bewertung zu rechtfertigen. So werden z.B. der Aufbau des Textes, die Angemessenheit der sprachlichen Mittel und die prinzipielle Sprachkompetenz bei der Verwendung analytischer Skalen stärker begutachtet als bei der Verwendung ganzheitlicher Skalen

(vgl. Barkaoui 2010: 66–68). Barkaoui (2011: 283–289) zeigt darüber hinaus, dass analytische Bewertungsskalen zu einer konsistenteren Bewertung der Texte führen. Eine solche analytische Bewertungsskala liegt auch im Falle der neu entwickelten Bewertungsmethode nach Kuiken und Vedder (2017: 323–325) vor. Die Bewertung der funktionalen Angemessenheit kann sowohl als eigenständige Bewertungsdimension als auch als Erweiterung der bisherigen CAF-Skala angesehen werden. Sie wird als aufgabenorientierte Bewertung verstanden, welche Produzenten und Rezipienten miteinbezieht. So stellt sie die Bewertung der Angemessenheit eines Textes in Bezug auf den gegebenen Kontext und den Rezipienten dar. Das FA-Bewertungstool berücksichtigt vier Textqualitätsdimensionen: Inhalt (*content*), Aufgabenerfüllung (*task requirements*), Verständlichkeit (*comprehensibility*) und Kohärenz und Kohäsion (*coherence and cohesion*) (vgl. Kuiken/Vedder 2017: 323–325). Die Qualitätsbewertung eines Textes auf der Basis der CAF Dimensionen verläuft nach Kuiken und Vedder (2017: 323–325) nicht unbedingt linear mit der Bewertung auf Basis der FA Dimensionen. Ob ein Text funktional angemessen ist, hängt also z.B. weniger von der Komplexität oder dem Wortschatzreichtum ab als vielmehr von der dem Kontext angemessenen Verwendung des Wortschatzes (vgl. Kuiken/Vedder 2017: 323–325). Ebenso wie Kuiken und Vedder (2017: 323–325), messen auch Crossley und McNamara (2010: 988) der Kohärenz und Kohäsion große Bedeutung bei. Sie analysierten die Korrelation zwischen der manuellen ganzheitlichen Bewertung und der manuellen Bewertung einzelner Textfaktoren, wie z.B. der Struktur, der Kohärenz und der Kohäsion des Textes. Resultierend aus der Analyse schreiben sie der Bewertung der Kohärenz eine Einflussnahme von 65 % auf die ganzheitliche Bewertung zu. Die analysierte Kohäsion korrelierte ebenfalls positiv mit der Textqualität, wurde jedoch nicht im Regressionsmodell berücksichtigt.

Die Bewertung von Kohärenz und Kohäsion innerhalb einer gemeinsamen Dimension in der FA-Skala von Kuiken und Vedder (2017) ist auf Basis von textlinguistischen Ansätzen nicht unumstritten. Schiftner (2013: 268–271) merkt darüber hinaus an, dass die Analyse der Kohärenz oft auf der Prämisse beruht, dass Kohärenz und Kohäsion in einem direkten Zusammenhang stehen. Sie stellt aus diesem Grund zwei Hypothesen auf: (1) Sollte keine Verbindung zwischen Kohärenz und Kohäsion bestehen, kann die Kohärenz eines Textes nicht durch Merkmale der Kohäsion bewertet werden. (2) Besteht eine Verbindung, kann die Kohärenz nicht allein durch Merkmale der Kohäsion bewertet werden, da das exakte Zusammenspiel der beiden Kategorien nicht eindeutig zu identifizieren ist. Die Kohärenz, welche sich auf das Zusammenspiel aus Konzepten und Relationen bezieht, sollte daher ergänzend im Hinblick auf die zugrunde liegenden Relationen innerhalb des Textes analysiert werden (vgl. Schiftner 2013: 268–271). Schiftner (2013: 271) schlägt dazu die Theorie rhetorischer Strukturen (engl.: *rhetorical structure theory*) (RST) vor,

die es erlaubt, unabhängig von der Oberflächenform Aussagen über den inneren Zusammenhang des Textes zu treffen. In der vorliegenden Studie wird diese Theorie mithilfe einer vertiefenden Analyse koreferenzieller Relationen auf Diskursebene angewandt, um zu eruieren, ob Abweichungen vom Modell auf rhetorische Strukturen zurückzuführen sind.

3 Zusammenhang zwischen manuellen Bewertungen und sprachlichen Merkmalen

Um Einsicht in die textuellen Merkmale, die ExpertInnen bei der Bewertung beeinflussen, zu erhalten, kann der Zusammenhang zwischen manuellen Bewertungen und automatisch ermittelten sprachlichen Textmerkmalen untersucht werden. Bisher wurde dieser Zusammenhang in Studien vor allem in Bezug auf das Englische als Lernersprache analysiert (vgl. Bestgen/Lories/Thewissen 2010; Crossley/McNamara 2010; Crossley/McNamara 2011; Crossley et al. 2016; Crossley 2020; Deane 2013; Guo et al. 2013; Tywoniw/Crossley 2019). So stellt Crossley (2020: 416) für englische Lernertexte z.B. fest, dass die Textlänge den stärksten Prädiktor von Textqualität und Schreibkompetenz darstellt (auch in Guo et al. 2013: 230; Jung et al. 2019: 47–48). Eine weitere Eigenschaft, die Einfluss auf die Bewertung der Schreibkompetenz nimmt, ist die lexikalische Gewandtheit. Diese Eigenschaft beruht nicht allein auf der Vorkommenshäufigkeit der verwendeten Wörter, sondern z.B. auch auf der Wahrscheinlichkeit diese in einem fortgeschrittenen Text anzutreffen und auf der Komplexität der Wörter selbst z.B. im Hinblick auf die Anzahl der Buchstaben oder Silben. Bezogen auf Lernertexte zeigt sich, dass Texte mit hoher Qualität eher seltene, weniger bildhafte und spezifischere Wörter enthalten (vgl. Crossley/McNamara 2011: 185; Crossley 2020: 418). Auf der Phrasenebene wird in kohärenteren Texten ein größeres Repertoire an Phrasentypen und -strukturen verwendet (vgl. Crossley 2020: 419). Ein weiterer Prädiktor der Schreibkompetenz ist die syntaktische Komplexität. So enthalten besser bewertete Texte komplexere, längere und diversere Phrasen in Kombination mit einer erhöhten Tiefe der Phrasenstruktur, z.B. im Hinblick auf subordinierende Phrasen (Casal/Lee 2019: 58; Crossley 2020: 423–424; Jung et al. 2019: 47; Lu 2017: 506).

Die Bewertungen der Merkmale Kohärenz und Kohäsion zeigen bei der Analyse der Korrelation zur manuellen Bewertung ein kontraintuitives Bild (vgl. Crossley/McNamara 2012: 130): Bezogen auf das Englische, enthält ein als kohärent bewerteter Text weniger explizite Kohäsionsmittel wie z.B. Konnektive und mehr implizite Mittel wie z.B. semantische Überlappungen (vgl. Crossley/McNamara 2010: 988–989, Crossley/McNamara 2011: 187; Crossley/McNamara 2012: 131–132; Crossley et al. 2016: 1234; Jung et al. 2019: 47). Crossley und McNamara

bezeichnen dies als *reverse cohesion effect* (Crossley/McNamara 2012: 130). Darüber hinaus zeigte sich eine positive Korrelation zwischen der semantischen Ähnlichkeit adjazenter Sätze bzw. Absätze und der Schreibkompetenz (vgl. Crossley et al. 2008: 1910; Crossley et al. 2016: 1234; Jung et al. 2019: 47). Dies deutet auf ein Potenzial der *Latent Semantic Analysis* (LSA) hin (vgl. Crossley et al. 2008:1910; Crossley/McNamara 2011: 179). Zudem sind in kohärenteren Texten weniger Überlappungen lexikalischer Elemente über Satzgrenzen hinweg zu finden. Sie zeigen eine erhöhte lexikalische Diversität auf (vgl. Bestgen et al. 2010: 392; Crossley/McNamara 2012: 130; Crossley 2020: 426). Die Ergebnisse der Komplexitäts- und Kohäsions- bzw. Kohärenzanalysen sind für die vorliegende Studie von Bedeutung, da sie einen Vergleich der englischen Tendenzen mit den zu analysierenden deutschen Tendenzen ermöglichen.

Im Vergleich zum Englischen ist die Forschungslage für DaF recht bescheiden. Bislang wurde nur in zwei Studien (Weiß/Riemenschneider/Schröter/Meurers 2019; Weiß/Meurers 2019) der Zusammenhang zwischen automatisch erhobenen sprachlichen Merkmalen und der Bewertung durch DaF-ExpertInnen untersucht. So analysieren Weiß und Meurers (2019: 431–432) den Zusammenhang zwischen Erwerbsniveaus auf Basis des *Common European Framework of Reference for Languages* (CEFR) und der Komplexität von Lernertexten. In Bezug auf Abiturtexte zeigen Weiß, Riemenschneider, Schröter und Meurers (2019: 38) auf, dass LehrerInnen bei der Bewertung einerseits auf quantitative Eigenschaften der Komplexität achten und sich andererseits auch stark durch die Korrektheit der Texte in Bezug auf Zeichensetzung, Rechtschreibung und Grammatik beeinflussen lassen. Eine DaF-Studie mit Fokus auf Kohärenz und Kohäsion gibt es bislang nicht.

Diese Forschungslücke für DaF soll die vorliegende Studie schließen. Die Forschungsfrage lautet dementsprechend: Wie hängt die Bewertung von Kohärenz und Kohäsion in narrativen DaF-Lernendertexten mit den quantifizierbaren sprachlichen Mitteln zusammen, die in den Texten verwendet wurden? Die vorliegende Studie greift die verwendete Methodologie der bestehenden Studien zum Englischen als L2 auf und untersucht, welche Merkmalskombination die Varianz der Kohärenz- und Kohäsionsbewertung mit höchster Wahrscheinlichkeit abbildet und ob sich ähnliche Tendenzen wie im Englischen feststellen lassen. Des Weiteren soll ein möglicher Zusammenhang der Bewertungen mit dem vorgegebenen Bewertungshorizont untersucht werden, methodologisch basiert dieser Teil der durchzuführenden Analyse dabei auf dem Vorgehen von Barkaoui (2010: 68), der diesen Zusammenhang für das Englische untersuchte.

4 Methode: Bewertung der Funktionalen Adäquatheit und Erfassung der sprachlichen Merkmale

Zur Erläuterung der Vorgehensweise werden in den folgenden Unterkapiteln das Korpus (4.1), die skalenbasierte Bewertung von Funktionaler Adäquatheit (4.2), die Erfassung der sprachlichen Merkmale (4.3) sowie die Annotation der Koreferenz (4.4) näher beschrieben.

4.1 Das Korpus

Das in dieser Studie zu analysierende Korpus setzt sich aus insgesamt 60 Texten, verfasst an zwei Testzeitpunkten, zusammen. Die Texte stammen von 30 StudentInnen der Angewandten Linguistik einer belgischen Universität, die ein Erasmus-Semester absolvierten und währenddessen an Deutschkursen teilnahmen. Alle TeilnehmerInnen sind MuttersprachlerInnen des Niederländischen und verfügten vor ihrer Abreise über Deutschkenntnisse auf dem ERK-Niveau B2-C1. Um Einsicht in die schriftliche Entwicklung der LernerInnen während des Erasmusaufenthalts zu erlangen, wurden die Texte vor und nach dem Auslandssemester verfasst. Die Aufgabenstellungen lauteten: „Auf dem Weg nach X: meine Erwartungen, meine Hoffnungen und meine Ängste“ (vor der Abreise) und „Meine Erasmus-Erfahrung: Was ich erwartet habe und wie es tatsächlich war“ (nach der Rückkehr). Die Texte wurden unter Testbedingungen handschriftlich und ohne Hilfsmittel verfasst. Die StudentInnen hatten 20 Minuten Zeit, die Texte zu verfassen (Strobl/Baten 2021: 16–17).

4.2 Skalenbasierte Bewertung der Funktionalen Adäquatheit

Die Bewertung der Texte (u.a. nach den Aspekten Kohärenz und Kohäsion) wurde von drei erfahrenen BewerterInnen und MuttersprachlerInnen des Deutschen mithilfe von Bewertungsskalen für funktionale Adäquatheit durchgeführt. Die an Kuiken und Vedder (2017: 335–336) angelehnten Bewertungsskalen bieten Deskriptoren für jeweils sechs Bewertungsstufen (1 = Text nicht kohärent und kohäsiv; 6 = Text stark kohärent und kohäsiv). Die Beschreibung der Entwicklung der hier angewandten Bewertungsskalen, welche ins Deutsche übersetzt und im Vergleich zu Kuiken und Vedder (2017: 335–336) um eine Skala reduziert wurden, findet sich in Strobl und Baten (im Druck). In dieser Studie werden nur die Bewertungen bezüglich Kohärenz und Kohäsion berücksichtigt (siehe Bewertungsskala im Anhang). Innerhalb der Deskriptoren für die einzelnen Notenstufen wurden die folgenden kohäsiven Mittel explizit genannt: anaphorische Sprachmittel (Stufen 1, 2, 3, 5 & 6), Wortwiederholungen (Stufen 1 & 5) und Konnektoren (Stufen 1 bis 6). In Fußnoten wird darüber hinaus auf folgende kohäsive Mittel hingewiesen:

Pronomina, Demonstrativa, Komparative, Pronominaladverbien und Synonyme (als Fußnote zu anaphorischen Sprachmitteln) sowie Konjunktionen, Adverbien, adverbiale Konstruktionen und Präposition mit Nominalisierung (als Fußnote zu Konnektoren).

4.3 Erfassung der sprachlichen Merkmale

Das Korpus wurde im Hinblick auf 196 größtenteils automatisch erfasste sprachliche Merkmale untersucht. Diese befanden sich auf Text-, Satz- sowie Wortebene und bezogen sich auf die Ausführlichkeit und die Variabilität der Form und der Bedeutung. Sie lassen sich in die folgend dargestellten Kategorien einteilen:

- **Lexikalische Komplexität:**
 - *Lexikalische Diversität und Varianz:* Diese Werte beziehen sich zumeist auf die Ergebnisse einer Analyse mithilfe des CTAP Tools (Chen/Meurers 2016) und umfassen verschiedene Arten der Berechnung der lexikalischen Dichte und Varianzen einzelner lexikalischer Wortarten.
 - *Kohäsion*
 - *Koreferenz:* Die mit CTAP berechneten Werte beschreiben die Anzahl und Variabilität der vorhandenen Überlappungen von Nomen, Argumenten und Stämmen bezogen auf globale und lokale Häufigkeiten. Globale Überlappung bezieht sich auf alle möglichen Satzpaare des Textes, während lokale Überlappung zwischen zwei adjazenten Sätzen berechnet wird.
 - *Konjunktionale Verknüpfungen:* Ebenfalls mit CTAP erhoben wurden die Häufigkeiten der Konnektive nach Breindl, Volodina und Waßner (2014) und Pasch, Brauße, Breindl und Waßner (2009). Sie werden nach Art der semantischen Relation (z.B. temporal) und Anzahl der Wörter (Ein- bzw. Mehrwort-Konnektive) unterteilt.
 - *Semantische Verknüpfungen:* Die von Weiß (2017: 76) als Relatiertheit beschriebene Kategorie bezieht sich auf die Anzahl der Synonyme, Hyponyme, Hyponyme und Wiederholungen innerhalb der Texte und basiert auf den in GermaNet (Hamp/Feldweg 1997; Henrich/Hinrichs 2010) vorhandenen Daten.
- **Syntaktische Komplexität:** Diese Kategorie bezieht sich auf die Analysen auf Phrasenlevel. Sie umfasst z.B. die Verteilung der Wortarten innerhalb eines Phrasentyps und die Varianz der Phrasentypen selbst. So wurde zum Beispiel analysiert, wie häufig subordinierende Nebensätze verwendet werden oder mit

welcher Häufigkeit Nomen vorliegen. Die Analyse wurde mithilfe von CTAP durchgeführt.

- **Morphologische Komplexität:** Die morphologische Komplexität, deren Analyse mit CTAP erfolgte, bezieht sich auf die Varianz der Verb- und Nomentypen (z.B. Infinitive, Vergangenheitsformen) sowie auf die unterschiedlichen Formen der Pronomen und Artikel (z.B. Verteilung der Pronomen in der ersten Person, Anzahl der indefiniten Artikel).
- **Flüssigkeit:** Diese Kategorie wurde anhand der Länge der Dokumente berechnet und umfasst die Anzahl der im Text zu findenden Worten. Da es sich um eine Schreibaufgabe mit zeitlicher Begrenzung handelt (20 Minuten), ist die Anzahl der Wörter aussagekräftig hinsichtlich der Flüssigkeit, mit der die Lernenden sich in der Fremdsprache äußern können.
- **Korrektheit:** Die Korrektheit wurde manuell berechnet. Sie beruht auf dem Quotienten der Anzahl der fehlerfreien Teilsätze und der Anzahl aller Teilsätze (*error-free clause ratio*) (vgl. Wolfe-Quintero et al. 1998: 33–62).

4.4 Analyseverfahren

Die Analyse beruht auf den Differenzen der Bewertungen der beiden Texte eines/einer jeden Lernenden hinsichtlich der Funktionalen Adäquatheit einerseits und hinsichtlich der sprachlichen Merkmale andererseits. Ein Wert von -2 für Funktionale Adäquatheit bedeutet z.B. eine deutliche Verschlechterung der Bewertung, während ein Wert von +1 eine leichte Verbesserung darstellt. Im Folgenden werden die verwendeten Analyseverfahren vorgestellt.

4.4.1 Korrelationsanalyse

Zur Eingrenzung der Menge der sprachlichen Merkmale, welche in die lineare Regression einbezogen werden, sowie zur Vermeidung von Multikollinearität der einzelnen Variablen wurde zunächst die Pearson-Korrelation berechnet. Dies geschah mithilfe von R 4.0.3 (R Core Team 2016). Der Schwerpunkt lag auf der Korrelation zwischen der Bewertung der Kohärenz (FA.C) und allen vorhandenen sprachlichen Merkmalen. Um eine Abhängigkeit der Merkmale untereinander auszuschließen, wurde zudem jeweils das am stärksten korrelierende Merkmal der zu findenden Gruppierungen ausgewählt. Anschließend wurden die Merkmale, welche keine signifikante Korrelation ($p < 0,05$) aufweisen, von der weiteren Analyse ausgeschlossen.

4.4.2 Multiple lineare Regression

Das Ziel der Analyse via multipler linearer Regression ist die Evaluation der einzelnen Merkmale und Merkmalskombinationen hinsichtlich ihres Einflusses auf die Bewertung der Kohärenz, welche durch MuttersprachlerInnen des Deutschen durchgeführt wurde. Dazu wurden die in der Analyse der Korrelation herausgestellten Merkmale sukzessiv in einem Modell auf ihren prädiktiven Wert überprüft. Um auch hier eine Abhängigkeit der Merkmale untereinander zu vermeiden, wurden der *variance inflation factor* (VIF) berechnet und die Prädiktoren auf einen VIF-Wert unter 2 eingegrenzt. Zwar zeigte die Analyse der *Cooks Distance* keine Auffälligkeiten ($< 0,05$), jedoch sind auf den Diagnostikplots (s. Anhang) Datenpunkte zu erkennen, die nicht die allgemeinen Tendenzen widerspiegeln (z.B. 3,7 und 15). Die dazugehörigen sechs Texte (jeweils Vor- und Nachtest dreier Studierender) bilden die Grundlage der Stichprobenanalyse. Auch hier beruht die Analyse auf den Differenzen der Bewertungen der beiden Texte (Vor- und Nachtest) eines/einer jeden Lernenden hinsichtlich der Funktionalen Adäquatheit einerseits und hinsichtlich der sprachlichen Merkmale andererseits.

4.5 Stichprobenanalyse der Ausreißer hinsichtlich Koreferenz

Auf Basis der Analysen für Korrelation und der multiplen linearen Regression wurden ‚Ausreißer‘ eruiert, d.h. Texte, bei denen das Verhältnis zwischen Qualitätsdifferenz bezüglich Funktionaler Adäquatheit einerseits und Differenz einzelner sprachlicher Merkmale andererseits nicht die im allgemeinen Modell ersichtlichen Tendenzen abbilden. Da Schiftner (2013: 268–271) bereits auf das Problem der Differenzierung zwischen Kohärenz und Kohäsion hingewiesen hat (siehe Kapitel 2.2), wird in der vorliegenden Studie die Koreferenz auf Diskursebene als potenzielle Schnittstelle zwischen den beiden Kategorien betrachtet und ihr möglicher Einfluss auf die Bewertung untersucht. Der Fokus auf die Diskursebene erlaubt Einblick in die Kohärenz, da sich nach de Beaugrande und Dressler (1981: 21) „die Kohärenz eines Textes [...] oft nur aus dem Blickwinkel des Gesamtdiskurses“ ergibt. Diese zusätzliche Analyse wurde hier als Pilotstudie aufgefasst, da eine Analyse aller 60 Texte hinsichtlich vorliegender Koreferenzketten aufgrund des Aufwandes durch eine manuelle Annotation und des noch nicht festgestellten Potenzi als Einflussgröße nicht sinnvoll erschien. Die Analyse der Stichprobe erfolgt deskriptiv und zeigt die individuellen Tendenzen der LernerInnen auf. Die folgenden Koreferenztypen werden in der weiteren Analyse berücksichtigt und beruhen auf den Annotationsrichtlinien von Reznicek (2013: 5–6):

- *Anaphoric* – Diskursreferent des Referenzausdrucks geht voraus
- *Cataphoric* – Diskursreferent des Referenzausdrucks steht nach

- *Coreferential* – Zwei referenzielle Ausdrücke referieren auf denselben Diskursreferenten, hängen jedoch nicht voneinander ab.
- *Bound* – externer Diskursreferent (bezieht sich vor allem auf Pronomen der ersten Person Singular und Plural)

5 Ergebnisse

Das folgende Kapitel stellt die Ergebnisse der bereits vorgestellten Analysemethoden vor. In Kapitel 5.1 werden die Ergebnisse der Pearson-Korrelation vorgestellt, in Kapitel 5.2 das lineare Modell und in Kapitel 5.3 die Koreferenzanalyse¹.

5.1 Pearson Korrelation

Es liegen 66 Korrelationen mit einer Korrelationsstärke von $r > +/- 0,25$ vor. 57 dieser Korrelationen sind positiv, allein 9 Merkmale korrelieren negativ mit FA.C, wie z.B. die Variabilität der Nomen und die Anzahl der Pronomen der ersten Person pro Satz. Darüber hinaus sind 19 Korrelationen signifikant ($p < 0,05$) (siehe Abbildung 1). Die stärksten positiven Korrelationen zu FA.C (Kohärenz und Kohäsion) finden sich bei einigen frequenzabhängigen Merkmalen, nämlich der Anzahl der lokalen Überlappungen der Stämme, der Anzahl an temporalen Konnektiven, der Anzahl der lokalen Nomenüberlappungen, der lexikalischen Diversität sowie der Gesamtanzahl der Wörter im Text. Weitere signifikant korrelierende Eigenschaften, die frequenzunabhängig sind, sind die durchschnittliche Anzahl an lokalen Überlappungen von Stämmen pro Satz sowie die durchschnittliche Anzahl an Hyponymen pro Token. Dies bedeutet, dass eine höhere Bewertung der Kohärenz und Kohäsion des Nachtests im Vergleich zum Vortest mit einer vermehrten Nutzung lokaler Nomen- und Stammüberlappungen, temporaler Konnektive und einer höheren lexikalischen Diversität zusammenhängt. Dies gilt ebenso für die Gesamtanzahl der Wörter, als auch für die Anzahl der Hyponyme pro Token. Darüber hinaus ist zu erkennen, dass der Wert für FA.A (Inhalt und Themenentfaltung) sowohl mit FA.C (Kohärenz und Kohäsion) als auch mit einigen frequenzabhängigen Kategorien (lexikalische Varianz und der Flüssigkeit) korreliert. Der Wert für FA.B (Verständlichkeit) korreliert mit keinem der mit FA.C korrelierenden Eigenschaften signifikant. FA.A und FA.C korrelieren miteinander signifikant ebenso wie FA.A und FA.B. Hieraus resultiert, dass die drei Konstrukte bzw. Bewertungskategorien zwar

¹ Im vorliegenden Artikel wird der sprachliche Zugewinn der Probanden während des Erasmusaufenthalt nicht thematisiert. Eine Diskussion dieses Zugewinns und der Einflussfaktoren findet sich in Strobl und Baten (2021).

zusammenhängen, jedoch unterschiedlich genug bewertet wurden, um sie unabhängig voneinander zu betrachten.

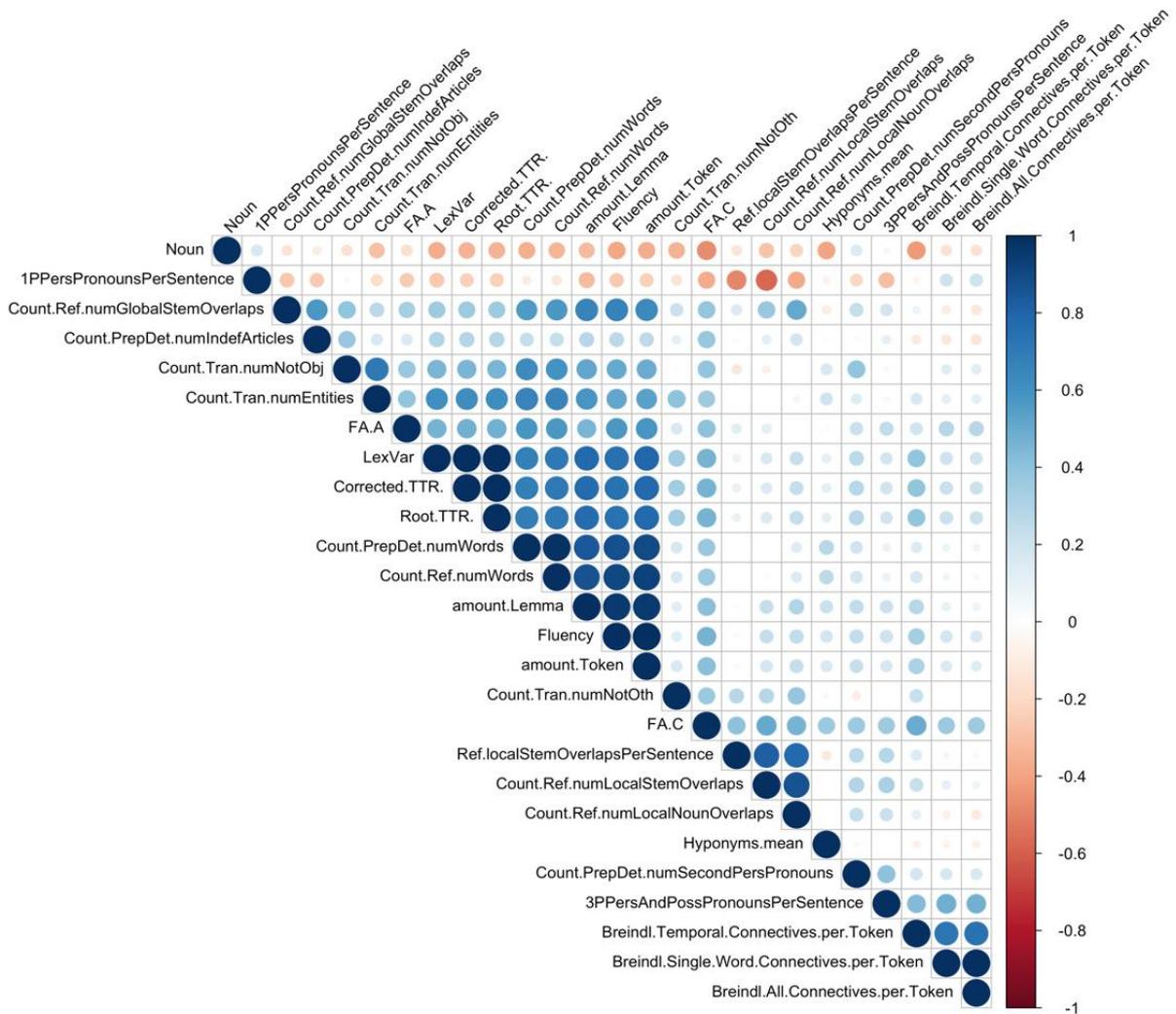


Abb. 1: Darstellungen der 25 signifikantesten Korrelationen (zu FA.C) mit einem Korrelationskoeffizienten zu FA.C von mindestens 0,25. Links und oben finden sich die Merkmale, rot (bis zu -1) bildet die negative Korrelation ab. Blaue Punkte bilden die Stärke der positiven Korrelationen ab. Die Namen beruhen auf gekürzten Versionen der von CTAP verwendeten Kategorien.

Tabelle 1 bildet die in Kapitel 4.1 beschriebenen bereinigten Korrelationen ab. So ist die Anzahl der lokalen Stammüberlappungen ein Merkmal der Koreferenz, die Type-Token-Relation nach Giraud ein Merkmal der lexikalischen Diversität, die temporalen Konnektive ein Merkmal der konjunkionalen Verknüpfungen und die Hyponyme ein Merkmal der semantischen Verknüpfungen. Die abgebildeten Werte zeigen ebenso wie Abbildung 1, dass eine höhere Bewertung der Kohärenz und Kohäsion des Nachttests im Vergleich zum Vortest mit einer vermehrten Nutzung lokaler Stammüberlappungen, temporaler Konnektive sowie Hyponyme und mit einer erhöhten lexikalischen Diversität und Flüssigkeit. Derselbe Zusammenhang gilt umgekehrt für niedrigere Bewertungen des Nachttests im Vergleich zum Vortest.

Tab. 1: Sprachliche Merkmale mit der stärksten Korrelation zu FA.C

Kategorie	Wert	Korrelation	p
Koreferenz	Lokale Stammüberlappungen pro Satz	0,402	0,030
Lexikalische Diversität	Type-Token-Ratio	0,469	0,010
Flüssigkeit	Flüssigkeit	0,463	0,011
Konnektoren	Anzahl temporaler Konnektive pro Token	0,500	0,006
Relatiertheit	Durchschnittliche Anzahl der Hyponyme pro Token	0,368	0,049
Ratings	FA.A	0,405	0,030
	FA.B	0,127	0,512
	FA.C	1	

5.2 Lineares Modell

Im Rahmen der weiteren Analyse wurden die im Modell berücksichtigten Kategorien differenziert. Die Anzahl der temporalen Konnektive könnte von der Aufgabenstellung beeinflusst werden, daher wurde diese Kategorie im Modell durch die übergeordnete Kategorie der Konnektive ersetzt, welche ebenfalls eine Korrelation von $r < 0,25$ aufzeigte. Da die Flüssigkeit und die lexikalische Varianz jeweils einen VIF-Wert von 2 aufweisen und somit miteinander korrelieren sowie Frequenzeffekte darstellen, wird ihre Einflussstärke getrennt vom Modell überprüft. Das Gesamtmodell ist statistisch signifikant ($F(3,25) = 7.571, p < .001$) und erklärt 42,32 % der Varianz der Bewertungen für Kohärenz und Kohäsion ($R^2 = .476$, adjusted $R^2 = .4132$). Der prädiktive Wert individueller Merkmale, d.h. der Einfluss auf das Gesamtmodell, wird aus Tabelle 2 (s. Spalte *R* und *adjusted R*) ersichtlich. Den stärksten Einfluss auf das Modell und somit auf die Bewertung der Kohärenz und Kohäsion hat dabei die Verteilung der lokalen Stammüberlappungen, gefolgt von den Hyponymen und den Konnektiven. Die Eigenschaften stellen ebenfalls die drei verschiedenen Formen der Kohäsion dar: Koreferenz (in Form von Stammüberlappungen), konjunktionale Verknüpfungen (in Form der Konnektive je Token) sowie semantische Verknüpfungen (in Form der Hyponyme). Die unabhängig voneinander getesteten Werte der lexikalischen Varianz und der Flüssigkeit finden sich in Tabelle 3.

Tab. 2: Prädiktoren des best-passenden linearen Regressionsmodells.

Wert	R	Adj. R	β	t	p	VIF
Anzahl Konnektive pro Token	0.127	0.094	3.045	2.601	0.015	1.007
Durchschnittliche Anzahl der Hyponyme pro Token	0.158	0.120	5.779	3.059	0.005	1.018
Lokale Stammüberlappungen pro Satz	0.191	0.161	1.759	3.020	0.006	1.013

Tab. 3: Prädiktive Werte der unabhängig berechneten linearen Regressionen

Wert	R	Adj. R	β	t	p
Type-Token-Ratio	0.220	0.191	0.495	2.761	0.010
Flüssigkeit	0.214	0.185	0.008	2.714	0.011

5.3 Analyse der Koreferenz

Die Analyse der Koreferenz wurde für die drei Ausnahmefälle jeweils individuell durchgeführt. Im Folgenden werden die Tendenzen pro ProbandIn aufgezeigt:

Tab. 4: Ergebnisse der Koreferenzanalyse der Texte von ProbandIn 3.

Statistik	Vortest	Nachtest
Bewertung FA.C	4	4
Anzahl der Sätze des Textes	10	21
Gesamtanzahl – anaphorische Relationen	3	8
Anzahl der Sätze mit anaphorischen Relationen	3	6
Gesamtanzahl – koreferenzielle Relationen	5	13
Anzahl der Sätze mit koreferenziellen Relationen	4	11
Gesamtanzahl – gebundene Relationen	19	30
Anzahl der Sätze mit gebundenen Relationen	9	13

Die Kohärenzbewertung von ProbandIn 3 (Tabelle 4) wurde sowohl im Vor- als auch im Nachtest mit 4 bewertet. Der Nachtest enthält mehr als doppelt so viele Sätze (21) wie der Vortest (10). Ebenso ungefähr verdoppelt hat sich die Verteilung der Koreferenzrelationen. Es finden sich pro Satz 1,25 (Vortest) und 1,15 (Nachtest) koreferenzielle Relationen, 1 (Vortest) bzw. 1,33 (Nachtest) anaphorische Relationen sowie 2,1 (Vortest) bzw. 2,3 (Nachtest) bound-Relationen.

Tab. 5: Ergebnisse der Koreferenzanalyse der Texte von ProbandIn 7.

Statistik	Vortest	Nachtest
Bewertung FA.C	3	5
Anzahl der Sätze des Textes	11	10
Gesamtanzahl – anaphorische Relationen	5	8
Anzahl der Sätze mit anaphorischen Relationen	4	6
Gesamtanzahl – koreferenzielle Relationen	6	8
Anzahl der Sätze mit koreferenziellen Relationen	3	7
Gesamtanzahl – gebundene Relationen	19	16
Anzahl der Sätze mit gebundenen Relationen	9	9

Die Texte von ProbandIn 7 (Tabelle 5) zeigen eine Verbesserung der Kohärenz von +2. Die Anzahl der Sätze bleibt unverändert. Es finden sich 1,25 (Vortest) und 1,33 (Nachtest) anaphorische Relationen pro Satz, 2,1 (Vortest) und 1,7 (Nachtest) bound-Relationen sowie 2 (Vortest) und 1,14 (Nachtest) koreferenzielle Relationen.

Tab. 6: Ergebnisse der Koreferenzanalyse der Texte von ProbandIn 15.

Statistik	Vortest	Nachtest
Bewertung FA.C	4	6
Anzahl der Sätze des Textes	16	16
Gesamtanzahl – anaphorische Relationen	1	10
Anzahl der Sätze mit anaphorischen Relationen	1	6
Gesamtanzahl – koreferenzielle Relationen	2	4
Anzahl der Sätze mit koreferenziellen Relationen	2	4
Gesamtanzahl – gebundene Relationen	18	23
Anzahl der Sätze mit gebundenen Relationen	13	12

Die Annotation der Verschriftlichungen von ProbandIn 15 (Tabelle 6) zeigen eine gleichbleibende Anzahl der Sätze sowie einen Anstieg der Bewertung um +2. Das Verhältnis zwischen den anaphorischen Relationen und der Anzahl der Sätze steigt von einer Relation pro Satz (Vortest) auf 1,66 Relationen (Nachtest). Die Verteilung der koreferenziellen Relationen ist gleichbleibend bei einer Relation pro Satz. Die Anzahl der Beziehungen insgesamt verdoppelt sich im Nachtest im Vergleich zum Vortest (2 → 4). Die bound-Beziehungen liegen im Durchschnitt 1,3 (Vortest) und 1,9 (Nachtest) Mal pro Satz vor.

6 Diskussion

Die Ergebnisse zeigen, welche sprachlichen Merkmale prädiktiven Wert für die skalenbasierte Kohärenz- und Kohäsionsbewertung haben, also die BewerterInnen beeinflussen. Im Folgenden werden die Ergebnisse der Analyse in den Gesamtzusammenhang der eingangs gestellten Forschungsfragen eingeordnet und im Licht der bisherigen Ergebnisse aus dem Bereich des Englischen besprochen.

In der vorliegenden Studie zeigt sich, dass die Textlänge bis 18,5 % der Varianz der Kohärenzbewertung beeinflusst. Je länger die Texte waren, desto besser wurde ihre Kohärenz und Kohäsion bewertet. Diese Tendenz findet sich auch in Bezug auf englische Lernertexte: die Textlänge stellt den stärksten Prädiktor von Textqualität und Schreibkompetenz dar (Crossley 2020: 416; Guo et al. 2013: 230–231; Jung et al. 2019: 47–48). Diese Einflussnahme in Kombination mit der Abhängigkeit vieler sprachlicher Merkmale von der Textlänge, die im Rahmen dieser Studie festgestellt wurden, könnte schlicht dahin gehend interpretiert werden, dass ein längerer Text auch mehr Potenzial für die Nutzung von kohäsiven Mitteln bietet. Als weiterer starker Prädiktor für die Kohärenzbewertung erwies sich in dieser Studie in Übereinstimmung mit Vorgängerstudien zum Englischen die lexikalische Diversität. Die Type-Token-Ratio erklärt bis zu 19 % der Varianz der manuellen Bewertungen für Kohärenz und Kohäsion.

Als kohärent bewertete Texte weisen viele semantische Verknüpfungen auf. Dies zeigt sich in der Anwesenheit der Merkmale Hyponyme und Stammüberlappungen im best-passenden Regressionsmodell. Auch hier zeigen englische Studien ähnliche Tendenzen. So haben bereits Crossley und McNamara (2010: 988–989) sowie Crossley und McNamara (2012: 130) die Hypothese aufgestellt, dass ein als kohärent bewerteter Text weniger *explizite* und mehr *implizite* Kohäsionsmittel enthält. Hinsichtlich der Gewichtung von konjunktionalen Verknüpfungen, ein typisches explizites Kohäsionsmittel, unterscheiden sich die Ergebnisse dieser Studie jedoch von bestehenden Studien zu englischen Lernertexten.

Bezüglich der Korrelation zwischen im Bewertungsraster aufgeführten kohäsiven Mitteln zeigt sich, dass Eigenschaften, die im Raster ausdrücklich genannt wurden, teilweise Bestandteil des ermittelten best-passenden Regressionsmodells für die manuellen Bewertungen sind. So werden die Merkmale anaphorische Mittel (z.B. in Form von Pronomina, Demonstrativa, Komparative, Pronomialadverbien sowie Synonyme) und Konnektoren (z. B. in Form von Konjunktionen, Adverbien, adverbiale Konstruktionen, Präposition mit Nominalisierung) als positive Faktoren für die Bewertung von Kohärenz und Kohäsion aufgezählt. Wortwiederholungen wurden als Negativbeispiel genannt. Die Ergebnisse der Studie zeigen, dass Konnektoren, Hyponyme und Stammüberlappungen Einfluss auf die Bewertung

genommen haben. Ein möglicher Zusammenhang findet sich bei den Konnektoren, aber auch bei Hyponymen als Form von anaphorischen Mitteln (vergleichbar zu Synonymen). Aus den Ergebnissen lässt sich insgesamt schließen, dass die Formulierungen im Bewertungsraster, wie in Vorgängerstudien (z.B. Barkaoui 2010: 68), einen deutlichen Einfluss auf die Bewertung durch die ExpertInnen ausgeübt hat.

Insgesamt lassen sich die Bewertungen der Texte nicht vollständig mithilfe der automatisch ermittelten Kategorien voraussagen. Die angeregt durch Schiftner (2013: 268–271) durchgeführte zusätzliche stichprobenartige Koreferenzanalyse zeigt unterschiedliche Tendenzen auf. Die Bewertung von ProbandIn 3 bleibt gleich. Sie zeigt zwar eine Verdopplung der Anzahl der Sätze, jedoch bleibt hierbei auch das Verhältnis zwischen der Anzahl der Sätze und den anaphorischen, koreferenziellen und bound-Beziehungen gleich. Die Verbesserung der Bewertung von +2 bei ProbandIn 7 und ProbandIn 15 bezieht sich auf zwei unterschiedliche Aspekte: In den Texten von ProbandIn 7 fallen besonders die Verteilung der koreferenziellen Relationen auf: Diese sinken von 2 Nennungen pro Satz im Vortest auf durchschnittlich 1,14 Nennungen im Nachtest. Da koreferenzielle Relationen sich vor allem in Form von Wiederholungen desselben expliziten Diskursreferenten finden, deutet dies auf eine Verringerung der vorhandenen Nomenwiederholungen hin. Bei ProbandIn 15 zeigt sich hingegen eine Erhöhung der anaphorischen Relationen im Nachtest. Während im schlechter bewerteten Vortest nur eine anaphorische Beziehung in einem der 14 Sätze vorlag, enthält der Nachtest insgesamt 10 anaphorische Koreferenzen, die auf 6 der 12 Sätze verteilt sind. Diese Ergebnisse deuten darauf hin, dass nicht nur die (automatisch erfassbare) Koreferenz zwischen adjazenten Sätzen, sondern auch die Koreferenz auf der Ebene des Gesamttextes die Kohärenzbewertung beeinflussen kann. Zwar finden sich diese Tendenzen in unterschiedlichen Formen sowie Stärken, jedoch weisen die Koreferenzanalysen der Vor- und Nachtests der drei näher analysierten ProbandInnen Schwankungen auf, die zu den skalenbasierten Bewertungen passen.

7 Fazit

Die vorliegende Pilotstudie, welche ein in DaF bislang unerforschtes Gebiet fokussiert, hatte zum Ziel eine mögliche Korrelation zwischen automatisch erhobenen sprachlichen Merkmalen und der skalenbasierten Bewertung von Kohärenz und Kohäsion durch ExpertInnen zu eruieren. Dies wurde mithilfe einer Korrelationsanalyse sowie eines linearen Regressionsmodells untersucht. Die Ergebnisse zeigen ähnliche Tendenzen beim Zusammenhang zwischen kohäsiven Mitteln und der Bewertung von englischen und deutschen Lernertexten: Je besser die Kohärenz und Kohäsion eines Textes bewertet wurde, umso weniger verwendeten die

ProbandInnen explizite kohäsive Mittel z.B. Wortwiederholungen und umso mehr verwendeten sie implizite kohäsive Mittel wie z.B. semantische Verknüpfungen. Eine Ausnahme bilden hierbei die konjunktionale Verknüpfungen, die im Deutschen einen stärkeren Einfluss auf die Kohäsionsbewertung zeigen als im Englischen. Darüber hinaus weisen die Ergebnisse darauf hin, dass die ExpertInnen sich durch die im Bewertungshorizont explizit genannten Kohäsionsmittel bei den Bewertungen beeinflussen ließen. Die analysierten sprachlichen Merkmale der Konnektive pro Token, der durchschnittlichen Anzahl der Hyponyme pro Token sowie der lokalen Stammüberlappungen pro Satz erklären 42 % der Variation in der Bewertung. Zudem zeigt sich eine Abhängigkeit des prädiktiven Wertes einiger frequenzabhängiger Kategorien von der Länge der Texte, welche ebenfalls einen prädiktiven Wert von über 20 % aufweist. Ein mögliches Argument für diese Korrelation ist, dass es in einem längeren Text mehr Möglichkeiten zur Nutzung von Kohäsionsmitteln gibt und der Text aus diesem Grund insgesamt kohäsiver erscheint.

Hier sollte erwähnt werden, dass die Aussagekraft der Studie durch die relativ geringe Anzahl der untersuchten Textpaare (30) eingeschränkt ist, wenngleich statistisch relevante Ergebnisse vorliegen. Darüber hinaus konnte der Einfluss von Koreferenz auf Textebene als potenziell relevanter Faktor im Rahmen dieser Studie nur stichprobenartig untersucht werden, da bislang kein automatisches Tool für das Deutsche existiert – im Englischen kann eine automatische Analyse von Koreferenz und LSA mithilfe von Coh-Metrix (Graesser et al. 2004; McNamara et al. 2010) durchgeführt werden. Allerdings hat die Stichprobe das Potenzial einer solchen Analyse auch für das Deutsche bestätigt.

Für die Bewertung von schriftlichen Leistungen im DaF-Unterricht ergeben sich aus der vorliegenden Studie folgende Implikationen:

- Bewertungsraster spielen eine große Rolle bei der Gewichtung vorhandener stilistischer Mittel. Sie sollten daher eine gezielte, am Erwartungshorizont angelehnte Auswahl an kohäsiven Mitteln konkret benennen z.B. „Kohäsive Texte enthalten mehr Konnektive wie z.B. Konjunktionen“. Dies erlaubt LernerInnen das Bewertungsraster zu kennen, sodass sie wissen, worauf sie achten können und wie sie bewertet werden.
- Der Einfluss der Textlänge und der lexikalischen Diversität deutet darauf hin, dass der Fremdsprachenunterricht nicht nur Wert auf die Korrektheit der Texte legen sollte, sondern auch die Flüssigkeit des Schreibens einüben sollte, da sie einen wichtigen Einfluss bei den Bewertungen in Testsituation hat.

Über diese Implikationen hinaus hat diese Studie einen großen Forschungsbedarf in Bezug auf die Analyse kohäsiver Mittel in deutschen Lernertexten aufgezeigt und zwar im Hinblick auf (a) die Entwicklung automatischer Analyseverfahren sowie (b) die Abgrenzung von Kohärenz und Kohäsion.

Der Forschungsbedarf bei der *Entwicklung automatischer Annotations- und Analysetools* der Kohäsion (a) zeigt sich vor allem im Unterschied der Anzahl und Nutzerfreundlichkeit der verfügbaren Tools des Englischen und des Deutschen. Hier bietet für das Deutsche allein CTAP eine nutzerfreundliche automatische Analyse eines gesamten Korpus an, während für das Englische eine Vielzahl an Tools existiert, mit denen unterschiedliche Merkmale in einer Benutzeroberfläche integriert untersucht werden können. Schließlich besteht Forschungsbedarf bei der getrennten Untersuchung von *Kohärenz und Kohäsion* (b). So hat Schiftner (2013: 268–271) aufgezeigt, dass diese Kategorien zwar im Wesen verknüpft sind, jedoch unterschiedliche Konstrukte beschreiben. Eine getrennte Analyse oder ein deutlicherer Rückbezug auf die Kohärenz, z.B. durch eine RST-Analyse, wären lohnende weitere Forschungsaufgaben.

Literatur

- Adamzik, Kirsten (2016): *Textlinguistik*. Berlin, Boston: De Gruyter.
<https://doi.org/10.1515/9783110339352>.
- Barkaoui, Khaled (2010): Variability in ESL essay rating processes: The role of the rating scale and rater experience. *Language Assessment Quarterly* 7: 1, 54–74. <https://doi.org/10.1080/15434300903464418>.
- Barkaoui, Khaled (2011): Effects of marking method and rater experience on ESL essay scores and rater performance. *Assessment in Education: Principles, Policy & Practice* 18: 3, 279–293.
<https://doi.org/10.1080/0969594X.2010.526585>.
- Bestgen, Yves; Lories, Guy & Thewissen, Jennifer (2010): Using latent semantic analysis to measure coherence in essays by foreign language learners? In: Bolasco Sergio; Chiari, Isabella & Giuliano, Luca (Hrsg.): *Proceedings of 10th International Conferences Journée d'Analyse statistique des Données Textuelles (JADT2010)*. Rom: LED, 385–395.
- Breindl, Eva; Volodina, Anna & Waßner, Ulrich (2014): *Handbuch der deutschen Konnektoren 2*. Berlin, München, Boston: De Gruyter.
<https://doi.org/10.1515/9783110341447>.

- Casal, J. Elliot & Lee, Joseph J. (2019): Syntactic complexity and writing quality in assessed first-year L2 writing. *Journal of Second Language Writing* 44, 51–62. <https://doi.org/10.1016/j.jslw.2019.03.005>
- Chen, Xiaboin & Meurers, Detmar (2016): CTAP: A web-based tool supporting automatic complexity analysis. In: *Proceedings of the workshop on computational linguistics for linguistic complexity (CLALC)*, 113–119. <https://aclanthology.org/W16-4113.pdf> (2.5.2021).
- Crossley, Scott (2020): Linguistic features in writing quality and development: An overview. *The Journal of Writing Research* 11, 415–443. <https://doi.org/10.17239/jowr-2020.11.03.01>.
- Crossley, Scott; Kyle, Kristopher & Dascalu, Mihai (2019): The Tool for the Automatic Analysis of Cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods* 51, 14–27. <https://doi.org/10.3758/s13428-018-1142-4>.
- Crossley, Scott; Kyle, Kristopher; Allen, Laura; Guo, Liang & McNamara, Danielle (2014): Linguistic Microfeatures to Predict L2 Writing Proficiency: A Case Study in Automated Writing Evaluation. *The Journal of Writing Assessment* 7: 1, 1–34. <http://www.journalofwritingassessment.org/article.php?article=74> (1.6.2021).
- Crossley, Scott; Kyle, Kristopher & McNamara, Danielle (2016): The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, 48, 1227–1237. <https://doi.org/10.3758/s13428-015-0651-7>.
- Crossley, Scott; McCarthy, Philip M.; Salsbury, Thomas & McNamara, Danielle (2008): LSA as a measure of coherence in second language natural discourse. *Proceedings of the 30th annual conference of the Cognitive Science Society*, 1906–1911.
- Crossley, Scott & McNamara, Danielle (2010): Cohesion, coherence, and expert evaluations of writing proficiency. *Proceedings of the Annual Meeting of the Cognitive Science Society* 32, 984–989.
- Crossley, Scott & McNamara, Danielle (2011): Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life-Long Learning* 21: 2-3, 170–191. <https://doi.org/10.1504/IJCEELL.2011.040197>.
- Crossley, Scott & McNamara, Danielle (2012): Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading* 35: 2, 115–135. <https://doi.org/10.1111/j.1467-9817.2010.01449.x>.
- De Beaugrande, Robert-Alain & Dressler, Wolfgang Ulrich (1981): *Einführung in die Textlinguistik* (Vol. 28). Tübingen: Max Niemeyer Verlag.

- Deane, Paul (2013): On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing* 18, 7–24.
<https://doi.org/10.1016/j.asw.2012.10.002>.
- Graesser, Arthur; McNamara, Danielle; Louwerse, Max & Cai, Zhiqiang (2004): Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers* 36, 193–202.
<https://doi.org/10.3758/BF03195564>.
- Guo, Liang; Crossley, Scott & McNamara, Danielle (2013): Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing* 18, 218–238. <https://doi.org/10.1016/j.asw.2013.05.002>.
- Hamp, Birgit & Feldweg, Helmut (1997): GermaNet - a Lexical-Semantic Net for German. *Proceedings of the ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*.
<https://aclanthology.org/W97-0802.pdf> (3.6.2021).
- Hancke, Julia (2013): *Automatic Prediction of CEFR Proficiency Levels Based on Linguistic Features of Learner Language*. Unveröffentlichte Masterarbeit. Eberhard Karls Universität Tübingen.
- Hancke, Julia; Vajjala, Sowmya & Meurers, Detmar (2012): Readability Classification for German using Lexical, Syntactic, and Morphological Features. *COLING*, 1063–1080. <https://aclanthology.org/C12-1065.pdf> (3.6.2021).
- Henrich, Verena & Hinrichs, Erhard (2010): GernEdiT - The GermaNet Editing Tool. *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010)*, 2228–2235. http://www.lrec-conf.org/proceedings/lrec2010/pdf/264_Paper.pdf (3.6.2021).
- Housen, Alex & Kuiken, Folkert (2009): Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics* 30: 4, 461–473.
<https://doi.org/10.1093/applin/amp048>.
- Jung, Yeon Joo; Crossley, Scott & McNamara, Danielle (2019): Predicting Second Language Writing Proficiency in Learner Texts Using Computational Tools. *The Journal of AsiaTEFL* 16, 37–52.
<https://doi.org/10.18823/asiatefl.2019.16.1.3.37>.
- Kuiken, Folkert & Vedder, Ineke (2014): Rating written performance: What do raters do and why? *Language Testing* 31: 3, 329–348.
<https://doi.org/10.1177/0265532214526174>.
- Kuiken, Folkert & Vedder, Ineke (2017): Functional adequacy in L2 writing: Towards a new rating scale. *Language Testing* 34: 3, 321–336.
<https://doi.org/10.1177/0265532216663991>.

- Kyle, Kristopher (2016): *Measuring Syntactic Development in L2 Writing: Fine Grained Indices of Syntactic Complexity and Usage-Based Indices of Syntactic Sophistication*. Dissertation, Georgia State University.
https://scholarworks.gsu.edu/cgi/viewcontent.cgi?article=1035&context=alesl_diss (5.6.2021).
- Kyle, Kristopher & Crossley, Scott (2015): Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application. *TESOL Quarterly* 49, 757–786. <https://doi.org/10.1002/tesq.194>.
- Kyle, Kristopher; Crossley, Scott & Berger, Cynthia (2018): The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior Research Methods* 50, 1030–1046.
<https://doi.org/10.3758/s13428-017-0924-4>.
- Kyle, Kristopher; Crossley, Scott & Jarvis, Scott (2021): Assessing the Validity of Lexical Diversity Indices Using Direct Judgements. *Language Assessment Quarterly* 18, 154–170. <https://doi.org/10.1080/15434303.2020.1844205>.
- Kunz, Kerstin & Lapshinova-Koltunski, Ekaterina (2014): Cohesive conjunctions in English and German: Systemic contrasts and textual differences. In: Vandelanotte, Lieven; Davidse, Kristin; Gentens, Caroline & Kimps, Ditte (Hrsg.): *Recent Advances in Corpus Linguistics: Developing and Exploiting Corpora*. Amsterdam/New York: Rodopi, 229–262.
https://doi.org/10.1163/9789401211130_012.
- Kunz, Kerstin & Lapshinova-Koltunski, Ekaterina (2015): Cross-linguistic analysis of discourse variation across registers. In: Ajmer, Karin & Hasselgard, Hilde (Hrsg.): *Special Issue of Nordic Journal of English Studies* Vol. 14, No. 1. Cross-linguistic Studies at the Interface between Lexis and Grammar, 258–288. <https://doi.org/10.35360/njes.347>.
- Link, Stephanie Maranda (2015): *Development and validation of an automated essay scoring engine to assess students' development across program levels*. Unveröffentlichte Dissertation, Iowa State University.
<https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=5574&context=etd> (1.8.2021).
- Lu, Xiaofei (2017): Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing* 34: 4, 493–511. <https://doi.org/10.1177/0265532217710675>.
- Lumley, Tom (2005): *Assessing second language writing: The rater's perspective*. Peter Lang.
- McNamara, Danielle, Crossley, Scott & McCarthy, Philip. M. (2010): Linguistic features of writing quality. *Written communication* 27: 1, 57–86.
<https://doi.org/10.1177/0741088309351547>.
- McNamara, Danielle; Louwerson, Max, McCarthy, Philip M. & Graesser, Arthur (2010): Coh-Metrix: Capturing Linguistic Features of Cohesion. *Discourse Processes* 47, 292–330. <https://doi.org/10.1080/01638530902959943>.

- Pasch, Renate; Brauße, Ursula; Breindl, Eva & Waßner, Ulrich (2009): *Handbuch der deutschen Konnektoren*. Berlin, New York: De Gruyter.
- R Core Team. (2016): R: A Language and Environment for Statistical Computing. Vienna, Austria.
- Reznicek, Marc (2013): *Linguistische Annotation von Nichtstandardvarietäten — Guidelines und „Best Practices“*. *Guidelines Koreferenz*. <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/nosta-d/nosta-d-cor-1.1/view> (1.6.2021).
- Schiftner, Barbara (2013): Analysing coherence in upper-intermediate learner writing. In: Díaz-Negrillo, Ana; Ballier, Nicolas & Thompson, Paul (Hrsg.): *Automatic treatment and analysis of learner corpus data*. Amsterdam: John Benjamins, 265–286. <https://doi.org/10.1075/scl.59.16sch>.
- Steiner, Erich. (2015): Contrastive studies of cohesion and their impact on our knowledge of translation (English-German). *Target* 27, 351–369. <https://doi.org/10.1075/target.27.3.02ste>.
- Strobl, Carola & Baten, Kristof (2021): Writing Development during Study Abroad: The Role of Language Contact and Social Networks. *Study abroad research in second language acquisition and international education* 6: 1, 124–62. <https://doi.org/10.1075/sar.19015.str>.
- Strobl, Carola & Baten, Kristof (im Druck): Assessing writing development during study abroad: The role of task and measures of linguistic and communicative performance. *TASK Journal on Task-Based Language Teaching and Learning*, 2 (1).
- Tywoniw, Rurik, & Crossley, Scott (2019): The effect of cohesive features in integrated and independent L2 writing quality and text classification. *Language Education & Assessment* 2: 3, 110–134. <https://doi.org/10.29140/lea.v2n3.151>.
- Vyatkina, Nina, & Housen, Alex (2020): Complexity. In: Paquot, Magali & Tracy-Ventura, Nicole (Hrsg.): *The Routledge Handbook of Second Language Acquisition and Corpora*. New York, Routledge, 318–331 <https://doi.org/10.4324/9781351137904-28>.
- Weiß, Zarah (2017): *Using Measures of Linguistic Complexity to Assess German L2 Proficiency in Learner Corpora under Consideration of Task-Effects*. Unveröffentlichte Masterarbeit, Eberhard Karls Universität Tübingen.
- Weiß, Zarah & Meurers, Detmar (2019): Broad linguistic modeling is beneficial for German L2 proficiency assessment. In: Abel, Andrea; Glaznieks, Aivars; Lyding, Verena & Nicolas, Lionel (Hrsg.): *Widening the Scope of Learner Corpus Research: Selected Papers from the 4th Learner Corpus Research Conference*. Louvain-La-Neuve: Presses Universitaires de Louvain, 419–435.

- Weiß, Zarah; Riemenschneider, Anja; Schröter, Pauline & Meurers, Detmar (2019): Computationally modeling the impact of task-appropriate language complexity and accuracy on human grading of German essays. *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, 30–45. <https://aclanthology.org/W19-4404.pdf> (5.6.2021).
- Wolfe-Quintero, Kate; Inagaki, Shunji; Kim, Hae-Young (1998): *Second language development in writing: Measures of fluency, accuracy, & complexity*. Manoa: University of Hawai'i, Second Language Teaching & Curriculum Center.

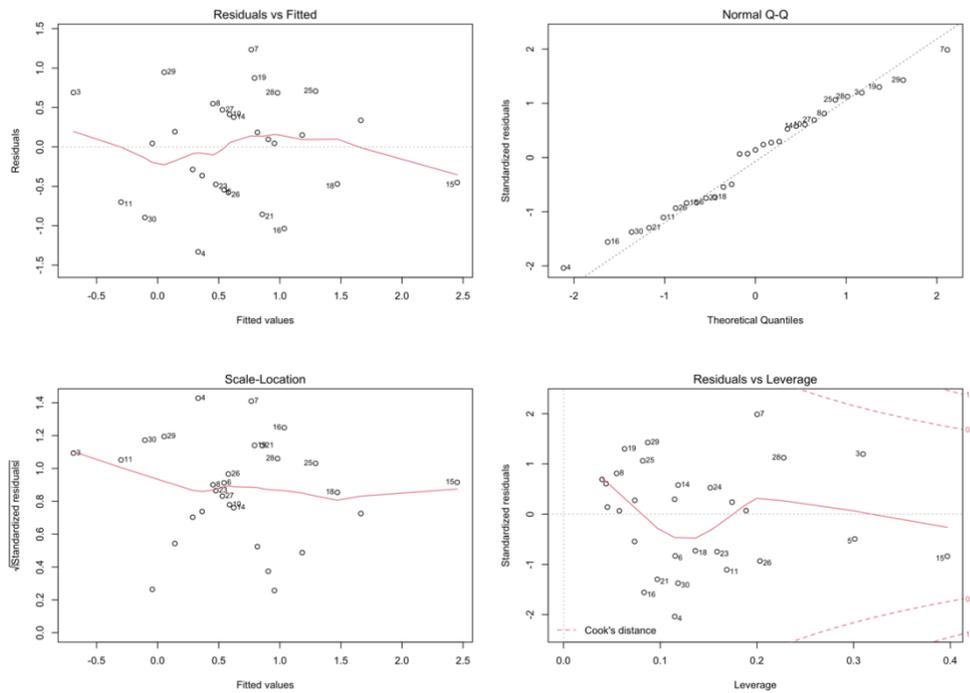
Anhang A: Bewertungsraster für Kohärenz und Kohäsion (Strobl/Baten 2021, in Anlehnung an Kuiken/Vedder 2017)

C. Kohärenz und Kohäsion	<p>Mit dieser Rubrik wird der Textzusammenhang, sowohl zwischen einzelnen Sätzen als auch zwischen Gedanken, bewertet. Im Zentrum steht dabei, ob der Gedankengang logisch aufgebaut und nachvollziehbar ist und ob diese Logik durch sprachliche Mittel gekennzeichnet ist. Leitfrage: Gibt es Gedankensprünge oder Sätze, die unverbunden nebeneinander stehen?</p>	
	1	<p>Der Text ist <i>überhaupt nicht kohärent</i>. Die Übergänge zwischen einzelnen Abschnitten / Gedanken sind völlig unzusammenhängend. Der Text ist <i>überhaupt nicht kohäsiv</i>. Der Schreiber benutzt keine anaphorischen Sprachmittel², bzw. nur Wortwiederholung, und so gut wie keine Konnektoren³.</p>
	2	<p>Der Text ist <i>kaum kohärent</i>. Die Übergänge zwischen einzelnen Abschnitten/ Gedanken sind oft unzusammenhängend. Gedanken werden wiederholt. Der Text ist <i>wenig kohäsiv</i>. Der Schreiber benutzt kaum anaphorische Sprachmittel und kaum Konnektoren.</p>
	3	<p>Der Text ist <i>einigermaßen kohärent</i>. Die Übergänge zwischen einzelnen Abschnitten/ Gedanken sind einigermaßen zusammenhängend, allerdings gibt es gedankliche Wiederholungen und eventuell Gedankensprünge. Der Text ist <i>einigermaßen kohäsiv</i>. Der Schreiber benutzt anaphorische Sprachmittel und Konnektoren, jedoch vor allem Konjunktionen.</p>
	4	<p>Der Text ist <i>kohärent</i>. Die Übergänge zwischen einzelnen Abschnitten/ Gedanken sind zusammenhängend, auch wenn der Schreiber manchmal Gedanken wiederholt. Der Text ist <i>kohäsiv</i>. Der Schreiber benutzt genügend Sprachmittel für interne Textverweise und zur Verdeutlichung von Zusammenhängen, und zwar nicht nur Konjunktionen.</p>
	5	<p>Der Text ist <i>sehr kohärent</i>. Wenn der Schreiber ein neues Thema einführt, geschieht dies in der Regel durch die Verwendung von Konnektoren oder Bindegliedern. Wiederholungen sind sehr selten. Der Text ist <i>sehr kohäsiv</i>. Anaphorische Sprachmittel sind zahlreich und die Gedanken sind durch eine Vielfalt von Konnektoren verbunden.</p>
	6	<p>Der Schreiber sorgt für <i>vollständige Kohärenz</i>. Die Übergänge zwischen einzelnen Abschnitten/ Gedanken geschieht durch die Verwendung von Konnektoren oder Bindegliedern. Der Text ist <i>äußerst kohäsiv</i>. Anaphorische Sprachmittel werden durchgängig benutzt. Konnektoren aller Art werden gewandt eingesetzt, um gedankliche Zusammenhänge zu verdeutlichen.</p>

² Anaphorische Sprachmittel verweisen auf ein vorgenanntes Element im Text zurück: Pronomina, Demonstrativa, Komparative, Pronominaladverbien, Synonyme, Oberbegriffe,...

³ Konnektoren verbinden Sätze / Satzteile logisch miteinander: Konjunktionen, Adverbien, Adverbiale Konstruktionen, Präposition mit Nominalisierung (z.B. "**aufgrund** der Zugverspätung (kam ich zu spät)...")

Anhang B: Diagnostikplot des berechneten linearen Modells.



Kurzbio:

Prof. Dr. Carola Strobl ist Professorin für Angewandte Sprachwissenschaft und Deutsch am Institut für Angewandte Sprachwissenschaft, Übersetzer und Dolmetscher der Universität Antwerpen. Ihre Forschungsinteressen umfassen Schreiberwerbsforschung, v.a. Entwicklung von Komplexität und Kohäsion, sowie computergestützte Fremdsprachen- und Übersetzungsdidaktik.

Helena Wedig M.A. ist Doktorandin der Angewandten Sprachwissenschaft am Institut für Angewandte Sprachwissenschaft, Übersetzer und Dolmetscher der Universität Antwerpen. Ihren Master in Linguistik mit Schwerpunkt Computerlinguistik hat sie an der Ruhr-Universität Bochum absolviert.

Anschrift:

Carola Strobl

Universität Antwerpen
 Stadscampus gebouw D
 Grote Kauwenberg 18
 B-2000 Antwerpen

Carola.Strobl@uantwerpen.be

Helena Wedig

Universität Antwerpen
 Stadscampus gebouw D
 Grote Kauwenberg 18
 B-2000 Antwerpen

Helena.Wedig@uantwerpen.be