https://bulletin.vals-asla.ch · e-ISSN: 2813-7973 · p-ISSN: 1023-2044

Karges, K., Studer, T. & Wiedenkeller, E. (2020). Textmerkmale als Indikatoren von Schreibkompetenz. *Bulletin suisse de linguistique appliquée* (Numéro spécial), 117-140. https://doi.org/10.26034/ne.vals-asla.2020.1220

Published in Bulletin VALS-ASLA spécial, 117-140, 2020,
which should be used for any reference to this work

Textmerkmale als Indikatoren von Schreibkompetenz

Katharina KARGES, Thomas STUDER & Eva WIEDENKELLER
Universität Freiburg
Departement für Mehrsprachigkeitsforschung und Fremdsprachendidaktik
Rue de Rome 1, 1700 Freiburg, Schweiz
katharina.karges@unifr.ch; thomas.studer@unifr.ch

In this paper, we will discuss the meaningfulness and productivity of certain learner text characteristics as measures of written language competence. Our analyses are based on a corpus of written L2 learner texts in English, French and German and L1 texts in French and German. All texts were elicited with the same set of 8 tasks, in Switzerland, from learners around the end of compulsory school.
One research interest during data collection was the effect of the different tasks on the learner texts' characteristics and quality. Thus, we will show how automatic measures such as text length, MTLD or measures of lexical sophistication perform to distinguish between different groups of texts in our corpus. Based on these, somewhat limited findings, we broaden our analysis to two more qualitative aspects of learner texts, which may yield more concrete evidence of language competence: cross-linguistic influence and lexical quality.

Keywords:
corpus linguistics, second language acquisition, communicative tasks, language competences.

Stichwörter:
Korpuslinguistik, Fremdsprachenerwerb, kommunikative Aufgaben, Sprachkompetenzen.

1. Einleitung

Eine zentrale Frage der Fremdsprachenforschung lautet, wie gut Lernende eine Fremdsprache zu einem bestimmten Zeitpunkt beherrschen, wie sie zu ihrer fremdsprachlichen Kompetenz kommen und wie sie diese weiterentwickeln können (vgl. z.B. Bleyhl 1996; Schneider 2007). Dazu werden in der Fachdidaktik und im schulischen Fremdsprachenunterricht laufend Annahmen getroffen und Empfehlungen formuliert. Dies geschieht jedoch in der Regel aufgrund von Erfahrungswissen, wohingegen einschlägiges empirisches Wissen erstaunlich lückenhaft ist. Um diese Lücken zu füllen, werden in der angewandten Linguistik Sprachproduktionen von Lernenden erhoben und auf die eine oder andere Weise analysiert. In der Korpuslinguistik liegt der Fokus dabei auf der Analyse von Merkmalen der Lernertexte, die Rückschlüsse auf die (produktive) Sprachkompetenz der Lernenden bzw. ihre Entwicklung zulassen. Eine wichtige Herausforderung stellt jedoch die Operationalisierung von Kompetenz dar, die Bestimmung ihrer Komponenten und die Auswahl relevanter Indikatoren in den entsprechenden Lernerproduktionen. Ausserdem bleibt die Frage, inwiefern einzelne Texte die Schreibkompetenz von Lernenden zuverlässig und vollständig abbilden können. Ein Faktor, der bei all dem eine grosse Rolle spielt, ist die Entstehung der Texte: Sowohl die Aufgabe, die der

Textproduktion zugrunde liegt, als auch weitere Entstehungsbedingungen (z.B. ob ein Text auf Papier oder am Computer geschrieben wurde) beeinflussen die Sprachproduktion.

Um diese Effekte näher zu untersuchen, variierten wir für die Datenerhebung unseres Lernerkorpus verschiedene Aufgabenmerkmale und Entstehungsbedingungen. Deren Analyse steht im Zentrum des vorliegenden Artikels, dessen Leitfrage wie folgt lautet: Inwiefern können Textmerkmale, die sich mittels quantitativer Methoden erfassen lassen, Auskunft über die fremdsprachliche Schreibkompetenz von Lernenden geben? Nach einer Definition von Schreibkompetenz stellen wir kurz und keineswegs vollständig die Diskussion um mögliche Textmerkmale dar, mit deren Hilfe Schreibkompetenz abgebildet werden könnte[1]. Besondere Aufmerksamkeit gilt dabei der Messung von Wortschatz und der Sprachmischung in den Texten der mehrsprachigen Lernenden. In den folgenden Abschnitten beschreiben wir das Korpus, unser Vorgehen bei der Analyse der Texte sowie ausgewählte Resultate. Eine Diskussion der Ergebnisse führt zu ersten Schlussfolgerungen und wirft neue Fragen auf.

2. Schriftliche Sprachkompetenz und ihre Messung

2.1 Schreibkompetenz messen

In Anlehnung an die Definition von Kompetenz, die den internationalen Vergleichsstudien der OECD entspringt (Weinert 2001), und die auch im Kompetenzbegriff des *Gemeinsamen Europäischen Referenzrahmens für Sprachen* (GER: Europarat 2001) zu finden ist, wird Schreiben in der aktuellen fremdsprachendidaktischen Diskussion etwa folgendermassen definiert: "mit Mitteln der schriftlichen Sprachproduktion zielführend zu handeln, Situationen erfolgreich zu bewältigen und das eigene kognitive (motivationale, affektive) System geeignet zu regulieren" (Grabowski et al. 2007: 45). Diese sehr breite Definition entspricht auch jener im GER, in dem Schreiben nicht als separate Kompetenz, sondern als eine Form (die schriftliche) von sprachlicher Produktion bzw. Interaktion aufgeführt ist, die Menschen "[bei der Ausführung von kommunikativen Aufgaben] strategisch planvoll einsetzen" (Europarat 2001: 21). In den entsprechenden Skalen und Deskriptoren der Referenzniveaus zeichnen sich höhere Kompetenzniveaus daher vor allem dadurch aus, dass neue Aufgaben hinzukommen und, ab B2, Aufgaben zunehmend effektiver, präziser und flexibler bearbeitet werden können (Council of Europe 2018: 68ff.; Europarat 2001: 66ff.). Um dieses Erfüllen von Aufgaben bzw. Bewältigen von Situationen für die Forschung und den Unterricht sichtbar(er) zu machen, werden seit geraumer Zeit Kompetenzmodelle

[1] Begrifflich verwenden wir im Folgenden 'Schreibkompetenz' und 'schriftliche Sprachkompetenz' synonym als Unterkonstrukt von Sprachkompetenz. Mit 'Textmerkmale' sind, sofern nicht anders vermerkt, Merkmale von Lernertexten gemeint.

entworfen, welche sich entweder auf kommunikative Kompetenz in ihrer Gesamtheit beziehen (Bachman & Palmer 2010; Canale & Swain 1980) oder einzelne Aspekte davon abbilden (Levelt 1993 zu einem Modell des Sprechens; Weir & Khalifa 2008 zur Definition von Lesekompetenz). Schreiben wird im englischsprachigen Raum meist prozessorientiert modelliert (v.a. Hayes & Flower 1980; Kellogg et al. 2013), was v.a. für die Schreibdidaktik eine grosse Rolle spielt. Diese Modelle fokussieren, wie der Name schon sagt, weniger auf das Produkt des Schreibens (den Text) als auf seinen Entstehungsprozess, weshalb sie es nur bedingt erlauben, aufgrund des Textes Rückschlüsse über die Schreibkompetenzen eines Individuums zu ziehen. Etwas klarer wird der Zusammenhang hingegen im Schreibkompetenzmodell von Becker-Mrotzek & Schindler (2007), welches im deutschsprachigen Raum diskutiert wird. Darin ist Schreibkompetenz ein mehrdimensionales Konstrukt, in dem deklaratives Wissen, Problemlöse-Wissen, prozedurales Wissen und metakognitives Wissen über verschiedene Aspekte des Schreibens hinweg zusammenspielen (ebd.: 12-16 und 24). Teile dieses Modells, v.a. Kenntnisse und Routinen in den Bereichen Orthografie, Lexik und Syntax, können so operationalisiert werden, dass aufgrund von Merkmalen des Textes auf vorhandene bzw. fehlende Schreibkompetenz in gewissen Bereichen geschlossen werden kann.

Ausgehend davon stellt sich dann die Frage, welche Textmerkmale operationalisiert und quantifiziert werden können, um Textqualität und Schreibkompetenz zu erfassen (Petersen 2019: 14; vgl. auch Harsch et al. 2007).

Petersen (2019: 14f.) unterscheidet dafür zunächst zwischen objektivierbaren Textmerkmalen, die automatisch erhoben oder von einer einzelnen Person festgestellt werden können und weniger eindeutig erfassbaren Merkmalen, die von zwei oder mehr Personen geratet werden müssen. In beiden Fällen gibt es eine Fülle von möglichen Variablen und Kriterien, deren Auswahl von der jeweiligen Fragestellung und der Definition von Schreibkompetenz abhängig ist. So werden für die Bewertung der Schreibkompetenzen einer bestimmten Zielgruppe, z.B. in Schulleistungsstudien, häufig holistische oder analytische Bewertungsskalen eingesetzt, die von Raterinnen und Ratern für jeden Text einzeln eingestuft werden (z.B. *Konsortium HarmoS Fremdsprachen* 2009; Harsch et al. 2007; vgl. auch Eberharter et al. 2018). Merkmale, die sich im Text eindeutig(er) identifizieren lassen, werden hingegen u.a. in der Forschung zum sog. *Task Based Learning*-Ansatz verwendet, um den Einfluss verschiedener Aufgaben oder Aufgabenmerkmale auf die Lernersprache zu untersuchen. Zentral ist dabei das CAF-Framework (*complexity, accuracy, fluency*; vgl. Housen & Kuiken 2009; Skehan 2009)[2], dessen drei Konstrukte z.T. sehr unterschiedlich operationalisiert werden: *Accuracy* (Korrektheit) wird meist

[2] Skehan fügt dem CAF-Framework in seiner Darstellung ein L wie Lexis (Wortschatz) hinzu. Dieses wird in Abschnitt 2.2 näher besprochen.

durch die Abweichung von einer Norm (d.h. Fehler) definiert, *fluency* (Flüssigkeit) durch verschiedene Masse, die die Geschwindigkeit der Produktion sowie die Menge, Position und/oder Art von Pausen beschreiben. *Complexity* wird einerseits in Bezug auf die Komplexität der Aufgabe und deren Auswirkungen und andererseits auf die der erzeugten Sprache definiert, wobei insb. Letzteres weiter unterteilt werden kann (Housen & Kuiken 2009: 462-464).

Ein stärker quantitativer Ansatz findet sich an der Schnittstelle zwischen Korpus- und Computerlinguistik. Hier wird eine Fülle von Textmerkmalen mit Computerunterstützung erfasst und statistisch ausgewertet (z.B. *Coh-Metrix*: McNamara et al. 2014; oder CTAP: Chen & Meurers 2016). Viele dieser Ansätze sind allerdings bislang nur für das Englische verfügbar und ihre Aussagekraft hinsichtlich der zugrundeliegenden Schreibkompetenz von Lernenden ist noch zu wenig erforscht.

2.2 Wortschatzqualität

Wortschatz (*lexis*, Skehan 2009) stellt eine unverzichtbare Voraussetzung für das Schreiben von Texten dar, allerdings ist der direkte Rückschluss von Wörtern in einem Text auf dessen Qualität oder gar auf die zugrundeliegende Kompetenz des Schreibenden sehr problematisch (Treffers-Daller et al. 2018: 306). Das hat u.a. damit zu tun, dass unklar ist, wie Wortschatz zuverlässig operationalisiert werden kann. Ein weit verbreiteter Ansatz hierfür ist lexikalische Vielfalt (*lexical diversity*, LD). Um diese zu erfassen, wurde zunächst mittels des Type-Token-Verhältnisses (*type token ratio*, TTR) der Anteil unterschiedlicher Wörter am Text festgestellt. Da das TTR mit steigender Textlänge aber immer geringer wird, wurden im Laufe der Jahre verschiedene verwandte Masse entwickelt, welche weniger von der Textlänge abhängig sein sollen (Olinghouse & Wilson 2013; Treffers-Daller et al. 2018). Angesichts des mässigen Erfolgs dieser Versuche legt Jarvis (2013: 18) jedoch dar, dass TTR und seine Ableitungen kaum auf ein zugrundeliegendes, theoretisch begründbares Konstrukt zurückgeführt werden können. Er schlägt stattdessen eine Reihe von Aspekten vor, die zusammen ein Lexical-Diversity-Konstrukt ausmachen könnten (ebd.: 22-25): Neben der Variabilität von Wortschatz, ausgedrückt durch ein LD-Mass, nennt er Wortmenge (*volume*), Gleichmässigkeit (*evenness*), Seltenheit (*rarity*), Verteilung (*dispersion*) und Vielfalt (*disparity*).

Was Jarvis *rarity* nennt, ähnelt in vielem dem Konzept des Wortschatzreichtums (*lexical sophistication/richness*, vgl. z.B. Kim et al. 2018). Jarvis schlägt vor, das Mittel der Ränge zu nutzen, welche die im Text vorkommenden Wörter in einem Referenzkorpus einnehmen (ebd.: 29f.). Andere Möglichkeiten, die vermutete Wortschatzbreite der Schreibenden zu erfassen, sind u.a. der Anteil "seltener" (oder gerade nicht seltener) Wörter am Text, das *Lexical Frequency Profile* (Vasylets et al. 2017), die Häufigkeit von Bi- und Trigrammen sowie Analysen von Hyponymie, Polysemie und Synonymie (Kyle & Crossley 2016: 14). Die

meisten dieser Analysen basieren auf einem Vergleich mit einem Referenzkorpus, allerdings tun sich auch hier eine Reihe von Fragen auf, insbesondere durch den nicht immer klaren Zusammenhang zwischen der Häufigkeit eines Worts, seiner Relevanz und seiner Lernbarkeit in verschiedenen L2-Kontexten.

Auch die Korrektheit des verwendeten Wortschatzes ist alles andere als leicht zu operationalisieren. Polio und Shea (2014) zeigen auf, wie die Abwesenheit von Fehlern[3] in verschiedenen Studien unterschiedlich gehandhabt wird. So gibt es neben holistischen Ratings durch Menschen auch objektivere Werte wie die Anzahl fehlerfreier T-Units oder Clauses, die Anzahl Fehler, den Anteil bestimmter Fehlertypen sowie mehr oder weniger 'schwerer' Fehler. All diese Masse und Herangehensweisen erfassen teilweise unterschiedliche Aspekte von Korrektheit und sind daher eher komplementär zu verstehen (ebda.: 22).

2.3 Crosslinguistic influence

Beim Aufbau und beim Gebrauch eines mehrsprachigen Repertoires ist wechselseitige Beeinflussung der beteiligten Sprachen hochgradig erwartbar (z.B. Jessner 2008), wobei Formen der Sprachmischung heute zunehmend als Indikator dafür interpretiert werden, dass die Lernenden die bereits vorhandenen sprachlichen und strategischen Wissensbestände für das Sprachenlernen nutzen. So weist z.B. der Begleitband des GER (Council of Europe 2018) Aspekte der Sprachmischung (auf tieferen Niveaus) und des Sprachwechsels (auf höheren Niveaus) explizit als Facetten mehrsprachiger Kompetenz aus (vgl. etwa die Skala "Auf einem mehrsprachigen Repertoire aufbauen", ebd.: 162).

Terminologisch ist dem mehrsprachigen Sprachgebrauch nicht leicht beizukommen. Der verbreitete Begriff *Codeswitching* etwa steht für eine Reihe von Ansätzen zur genaueren linguistischen Bestimmung des Zusammenspiels von Sprachen bei der zwei- und mehrsprachigen Sprachverwendung (vgl. für eine Systematik Treffers-Daller 2009). Für die vorliegenden Zwecke verwenden wir den neutraleren Oberbegriff *crosslinguistic influence* (CLI). CLI wird gemäss Odlin (2012) grob synonym zu Transfer verwendet und steht somit sowohl für erleichternde als auch für erschwerende Effekte (positiver vs. negativer Transfer/Interferenz), die sich infolge von Kontrasten zwischen den Sprachen eines individuellen Repertoires und aktuell gelernten Sprachen ergeben können.

Was im Begleitband des GER als Kompetenz beschrieben wird, ist primär als sprachenpolitisch motivierte Soll-Vorstellung zu sehen, die auf skalierten Projektionen von Sprachexpertinnen und -experten basiert, nicht auf der Untersuchung von Lernerperformanzen (vgl. Bärenfänger et al. 2019;

[3] 'Fehler' wird innerhalb einer Sprachgemeinschaft als relativ konsensfähig angenommen (Polio & Shea 2014: 10).

Studer i. V.). Sucht man nach Konstrukten, welche entsprechende Untersuchungen im Bereich der zwei- und mehrsprachigen Fähigkeiten und Fertigkeiten leiten könnten, trifft man oft auf allgemeine und vage Konzeptionen. Treffers-Daller (2018) beispielsweise rekurriert auf Bachman & Palmer (2010), übernimmt die dort modellierten linguistischen, soziolinguistischen und strategischen Fähigkeiten auch für Zwei- und Mehrsprachige und macht geltend, dass diese Fähigkeiten bei Letzteren im Plural zu konzipieren seien. Eine Folge solch allgemeiner Konstrukte ist eine Vielzahl unterschiedlicher Operationa-lisierungen und methodischer Zugänge.

Geht es um CLI im engeren Sinn, konkurrieren auf der theoretischen Ebene zur Zeit vier Positionen, bei denen der Erwerb morphosyntaktischer Strukturen in einer L3 bzw. einer weiteren L2 im Vordergrund steht (Efeoglu et al. 2019; Slabakova 2017). Als Grundhypothese gilt der L1-Faktor, wonach die Erstsprache als eine Art Filter wirkt, der bestimmt, welche L2-Merkmale für den L3-Erwerb transferiert werden können. Dieser Position entgegengesetzt ist der L2-Status-Faktor (vormals auch *foreign language effect*, vgl. Ecke 2015: 147, mit Verweis auf Meisel 1983): Demnach ist für den L3-Erwerb nicht die L1, sondern die L2 entscheidend, was mit ihrer kognitiven Prominenz und (gedächtnis-)psychologischen 'Nähe' zur Verarbeitung der L3 begründet wird. Die beiden weiteren Modelle gehen schliesslich davon aus, dass alle zuvor gelernten Sprachen als Quelle für CLI in Frage kommen. Beim *Cumulative Enhancement Model* können sprachliche Erfahrungen den L3-Erwerb begünstigen oder diesem gegenüber neutral bleiben; negativen Transfer gibt es in diesem Modell nicht. Demgegenüber betont das *Typological Primacy Model* die Bedeutung der strukturellen Ähnlichkeit zwischen den Grammatiken der beteiligten Sprachen. Ausschlaggebend für Transfer ist hier der bestehende syntaktische Verarbeitungsmechanismus (Parser), der die neue Sprache tentativ, d.h. auf Basis wahrgenommener Ähnlichkeiten verarbeitet, was auch zu negativem Transfer führen kann[4].

Eine offene Frage in der Diskussion um diese Positionen ist, ob der oft beobachtete Einfluss einer nahverwandten L2 auf die L3 eine Folge typologischer Ähnlichkeit ist, mehr mit dem L2-Status-Faktor zu tun hat oder aber am besten mit einer Kombination von beidem (und weiteren Faktoren) erklärt werden sollte (Ecke 2015: 147ff.), wobei 'weitere Faktoren' lernerseitige Variablen (z.B. Kompetenzen in jeder Sprache, Erwerbsreihenfolge, Kontaktdauer), sprachliche Variablen (Komplexität zielsprachlicher Strukturen

[4] In diesem Modell steht "typologisch" für strukturelle Ähnlichkeiten lexikalischer oder grammatischer Einheiten auf der Ebene mentaler Repräsentationen und wird von "psychotypologisch" abgegrenzt (Slabakova 2017: 663, Fussnote 2). Komplizierend ist dabei, dass typologische und psychotypologische Aspekte (subjektive Wahrnehmungen von Ähnlichkeiten und Unterschieden zwischen Sprachen) zwar nicht immer, aber oft zusammenfallen (Ecke 2015: 147, mit Verweis u.a. auf Singleton 2012).

und spezifische Wortcharakteristika wie Frequenz) und situative Variablen (u.a. Aufgaben und Gesprächspartner) umfassen.

3. Lernertexte im SWIKO-Korpus

Die im Folgenden besprochenen Lernertexte entstammen der ersten Aufbauphase des Schweizer Lernerkorpus (SWIKO). Sie wurden zu zwei Zeitpunkten in der Westschweiz und in der Deutschschweiz in regulären Schulklassen erhoben. Während der Erhebung[5] bearbeiteten die Schülerinnen und Schüler acht verschiedene Aufgaben in je zwei der drei ihnen zur Verfügung stehenden Sprachen (Schulsprache + zwei Schulfremdsprachen), davon vier am Computer in einer Sprache und vier weitere auf Papier in einer anderen Sprache. Die Aufgabenstellung wurde jeweils in der Schulsprache gestellt, unabhängig davon, in welcher Sprache der Zieltext zu schreiben war (vgl. hierzu Karges et al. i. V.). Nach der Aufbereitung der Texte (s. Abschnitt 3.1) entstand so ein Korpus, welches aktuell 1452 Texte mit 88'554 Tokens enthält (Näheres dazu in Tabelle 1).

Textsprache	Deutsch		Französisch		Englisch
Status	SchS	1. FS	SchS	1. FS	2. FS
Anzahl Lernende	90	45	41	99	138 d: 99, f: 39
Klassenstufen	11H & 1. MatS	10H	10H	11H & 1. MatS	10H (f), 11H & 1. MatS (d)
Anzahl Texte	329	162	144	326	491 d: 348, f: 143
Anzahl Tokens	22'602	6'140	9'971	17'060	32'781 d: 26'114, f: 6'667

Tabelle 1: Übersicht über das SWIKO-Korpus, Stand: 14.08.2019[6].

Wie Tabelle 1 zu entnehmen ist, unterscheiden sich die Lernenden in den beiden Sprachregionen zum Zeitpunkt der Datenerhebung insofern voneinander, dass die Lernenden in der Westschweiz (10H) jünger waren als jene in der Deutschschweiz (11H bzw. 1. MatS). Ausserdem sei darauf hingewiesen, dass die teilnehmenden Klassen in der Westschweiz jeweils aus dem mittleren Leistungsniveau der Sekundarschulen stammten, während in der

5 Neben der Erhebung der schriftlichen Texte wurden von einzelnen Lernenden auch mündliche Texte erhoben. Diese werden hier nicht besprochen.

6 SchS: Schulsprache, FS: Fremdsprache, d: Schulsprache Deutsch, f: Schulsprache Französisch, H: Klassenstufe gemäss HarmoS, MatS: Maturitätsschule.

Deutschschweiz auch Lernende aus dem höchsten Leistungsniveau teilgenommen haben.

Aus organisatorischen Gründen liegen von den Lernenden aus der Westschweiz leider keine (sprach-)biographischen Informationen (z.B. allfällige weitere Sprachkompetenzen) vor, weshalb in diesem Artikel auf eine nähere Betrachtung dieser Aspekte verzichtet werden muss.

3.1 Transkription und Annotation

Alle Lernertexte wurden von einem Team speziell trainierter Hilfskräfte originalgetreu transkribiert und annotiert[7]. Dabei wurden u.a. orthografische Fehler und ihre Korrektur erfasst sowie Wörter und Wortgruppen markiert, die nicht in der Sprache des Textes geschrieben wurden. Auch gelöschte oder nachträglich hinzugefügte Elemente (bei handgeschriebenen Texten) oder z.B. Kommentare, die sich nicht auf die Aufgabe bezogen, wurden vermerkt. Elemente, die Rückschlüsse auf den Autor oder die Autorin zuliessen, wurden anonymisiert. So entstanden Transkripte, die die Originaltexte im Detail abbilden. Durch ein *Part-of-Speech*-Tagging ergänzt (POS-Tagging mit Treetagger; Schmid 2018), können sie nun in verschiedenen Dateiformaten auf der Datenbank SWIKOweb[8] abgerufen werden.

3.2 Analysen

Im Zuge des POS-Taggings wurden für jeden Text zahlreiche Zusatzinformationen erfasst, darunter die Textlänge, die Anteile der Wortarten im Text, Anzahl und Länge der Sätze und verschiedene Masse lexikalischer Diversität (unter Verwendung von R und v.a. des Packages "koRpus": Michalke 2017; R Core Team 2017). Ausserdem sind die Transkripte u.a. in Tabellenform verfügbar, was diverse Wortschatzanalysen sowie das Auffinden und Analysieren von Annotationen möglich macht.

3.2.1 Lexikalische Masse

Textlänge wird als Anzahl aller Tokens ausgedrückt, die während des POS-Taggings erfasst wurden. Dazu gehören sämtliche Wörter und Abkürzungen im Text (z.B. "TV"), auch jene in anderen Sprachen. Unter Ausschluss der nicht-zielsprachlichen Wörter wurden drei Masse *lexikalischer Diversität* automatisch berechnet: *Carroll's corrected TTR* (CTTR)*,* HD-D sowie MTLD (Michalke 2019; vgl. für eine Besprechung der Masse McCarthy & Jarvis 2010).

---

7 Das Transkriptions- und Annotationssystem basiert auf einem XML-Skript, welches uns freundlicherweise vom Team des MERLIN-Projektes zur Verfügung gestellt wurde: "MERLIN - Mehrsprachige Plattform für die Europäischen Referenzniveaus: Untersuchung von Lernersprache im Kontext" (Projektnummer: 518989-LLP-1-2011-1-DE-KA2-KA2MP).

8 Zugang zu dieser Online-Datenbank wird auf Anfrage gewährt. Auskunft gibt das Institut für Mehrsprachigkeit der PH Freiburg und der Universität Freiburg/Schweiz.

Für eine grobe Erfassung der *Korrektheit* eines Textes wurden die Anzahl und der Anteil jener Wörter erfasst, die während der Annotation orthographisch korrigiert wurden. In einem zweiten Schritt wurde diesem Fehlerindex die Anzahl von Wörtern hinzugefügt, die nicht in der Sprache des Textes geschrieben worden waren. So werden drei potenzielle Fehlerquellen erfasst: a) Wörter, die nicht zielsprachengerecht geschrieben wurden (was als ein Teil von Wortschatzkenntnis interpretiert werden kann, z.B. **Familli*), b) Wörter, die einer anderen Wortart angehören als im aktuellen Kontext möglich (was auf fehlerhafte Wortwahl zurückzuführen sein kann, z.B. *ich *sprache* anstelle von *ich spreche*) und c) Wörter, die nicht in der Zielsprache verwendet wurden (was auf Lücken im Wortschatz hinweist, z.B. *parce que* in einem deutschen Text). Obwohl so klar nur ein Teil von dem erfasst wird, was die Korrektheit eines Textes ausmacht, vermögen die beiden so entstehenden Fehlerindizes zumindest einen Eindruck davon zu vermitteln, wie zielsprachlich die Lernenden die Texte geschrieben haben.

Um schliesslich die Wortschatzbreite in den Texten zu erfassen, wurden für jede der drei Textsprachen Referenzkorpora hinzugezogen[9], um den Anteil der häufigsten 1000 Tokens am Text auszugeben und den mittleren Rang der lemmatisierten Tokens in den Referenzkorpora zu erfassen (wie in Jarvis 2013: 29f. vorgeschlagen).

3.2.2 Wortschatzqualität

Um die Verwendung des Wortschatzes näher analysieren zu können, wurden die Texte in sechs Subkorpora (Schulsprache Deutsch, Fremdsprache Deutsch, Schulsprache Französisch, Fremdsprache Französisch, Fremdsprache Englisch mit Schulsprache Deutsch, Fremdsprache Englisch mit Schulsprache Französisch) unterteilt. Für jede Gruppe wurden dann die häufigsten Tokens bzw. Lemmata identifiziert. Ausserdem wurden die Subkorpora in den jeweiligen Sprachen untereinander verglichen, indem jene Tokens bzw. Lemmas identifiziert wurden, die in einem Korpus vergleichsweise häufiger vorkommen als im anderen (sog. *Keywords*[10]). Bei diesen Analysen

[9] Englisch: COCA, mithilfe einer frei zugänglichen Liste der 5000 häufigsten Lemmas (https://www.wordfrequency.info/top5000.asp); Französisch: Lexique 3.81 (http://www.lexique.org/); Deutsch: DeReKo 2014 II des Instituts für Deutsche Sprache (http://www1.ids-mannheim.de/kl/projekte/methoden/derewo.html).

[10] Keywords wurden folgendermassen berechnet: Nach einer Bereinigung der Texte wurde die Häufigkeit aller Wörter bestimmt und auf 1000 Wörter standardisiert, wobei Wörter, die nur in einem der beiden Subkorpora auftauchen, im anderen einen Wert knapp über Null erhielten, um nicht aus der Analyse ausgeschlossen zu werden. Wörter, die in beiden Subkorpora nur einmal oder seltener vorkamen, wurden von der Analyse ausgeschlossen. Die so entstandenen Häufigkeitswerte in den beiden Subkorpora wurden dann durcheinander dividiert. Je nach Divisor weist dann ein besonders hoher oder ein besonders niedriger *Keywordness*-Wert auf verhältnismässig häufiges Auftreten hin oder umgekehrt. Die *Keywordness*-Analyse wurde getrennt für Tokens und Lemmata durchgeführt.

wurden Homographen nicht unterschieden. Alle Analysen wurden über alle Aufgaben hinweg und für jede Aufgabe einzeln durchgeführt.

3.2.3 Crosslinguistic influence

Wie bereits erwähnt wurden nicht-zielsprachliche Wörter und Wortgruppen im Zuge der Transkription als solche markiert. Deutsche, französische und englische Wörter wurden auch nach Sprache unterschieden. Generell wurden nur Wörter und Wortgruppen markiert, wenn sie eindeutig nicht oder nicht in der vermuteten Bedeutung zum Wortschatz der Zielsprache gehörten. Im Zweifel wurden einschlägige Wörterbücher konsultiert und es wurde tendenziell im Sinne des Lernenden (d.h. gegen eine Markierung als nicht-zielsprachlich) entschieden. Alle Analysen zu CLI basieren auf der Identifikation dieser Tags, wobei nicht unterschieden werden kann, ob den Lernenden die Verwendung des nicht-zielsprachlichen Begriffes bewusst war oder nicht.

4. Resultate

Durch die eher geringe Anzahl der Texte, die generell kurzen Texte und die grosse individuelle Streuung sind alle im Folgenden gemachten Aussagen nicht im statistischen Sinne zu verstehen, sondern eher als Tendenzen (vgl. dazu auch die Abbildungen). Aus diesem Grund wird auf die Angabe von statistischen Signifikanzen generell verzichtet.

4.1 Textlänge

In einer früheren Phase des SWIKO-Projektes, in der nur die Texte aus der Westschweiz vorlagen, stellten wir Unterschiede in der Textlänge zwischen verschiedenen Aufgaben fest (Karges et al. 2019: 152ff.). Die französischsprachigen Lernenden hatten in ihren Fremdsprachen, vor allem in Deutsch, eher kürzere Texte geschrieben, wenn sie argumentieren mussten oder ein schulisches Thema gefordert war. Hingegen waren die Texte länger, wenn das Thema ihrer Lebenswelt näher war. Dieses Muster entsprach unseren Erwartungen aus dem Studium der Literatur (ebd.: 143f.). In den später erhobenen Texten aus der Deutschschweiz fanden wir diese Effekte hingegen praktisch nicht mehr: Die deutschsprachigen Schülerinnen und Schüler schrieben in ihren beiden Fremdsprachen in allen Aufgaben ähnlich lange Texte.

Eine vergleichbare Beobachtung machten wir in Bezug auf die Länge der am Computer und auf Papier geschriebenen Texte: Hatten die französisch-sprachigen Lernenden vor allem in Deutsch am Computer noch kürzere Texte geschrieben, stellten wir bei den neuen Texten aus der Deutschschweiz das Gegenteil fest: Texte in französischer Sprache waren in beiden Fällen ähnlich lang, Texte auf Englisch am Computer sogar im Schnitt länger (vgl. Abbildung 1).

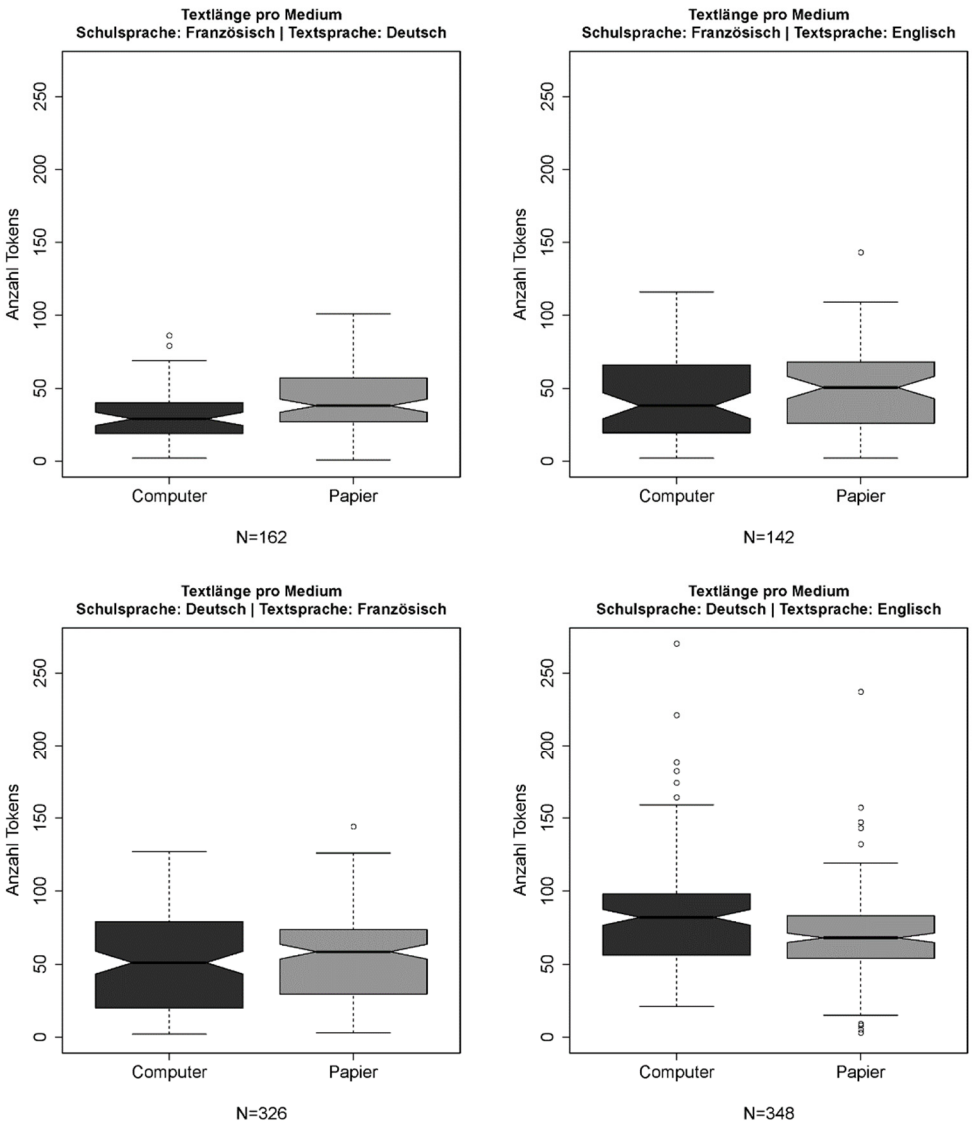

Abbildung 1: Textlänge am Computer und auf Papier, nach Sprachkombination.

Ein genauer Blick auf Abbildung 1 weist auch auf einen weiteren Unterschied zwischen den beiden Subkorpora hin: Die Schülerinnen und Schüler in der Deutschschweiz schrieben sichtlich längere Texte in Englisch als ihre Pendants in der Westschweiz. Auch in der je ersten Fremdsprache der beiden Sprachgruppen ist diesbezüglich ein Unterschied zu beobachten, er ist aber geringer und könnte auch darauf zurückzuführen sein, dass französische Texte bei gleichem Inhalt grundsätzlich länger sind als deutsche.

4.2 Wortschatz

4.2.1 Wortschatzvielfalt

Die grafische Kontrolle der Verteilung der LD-Masse zeigte, dass über alle Texte hinweg nur MTLD eine ausreichend starke Streuung aufweist, dass es

als von der Textlänge weitgehend unabhängig eingestuft werden kann[11]. Aus diesem Grund werden im Folgenden nur Gruppenunterschiede aufgezeigt, die sich aufgrund der MTLD-Werte zeigen.

Bezüglich der verschiedenen Sprachkombinationen lassen sich vor allem zwei Beobachtungen festhalten: Zum einen haben die Texte in der Schulsprache deutlich höhere MTLD-Werte als die jeweiligen fremdsprachlichen Texte. Sie weisen auch eine grössere Streuung der Werte auf. Zum anderen unterscheiden sich die MTLD-Werte der Texte in den beiden Fremdsprachen der jeweiligen Sprachgruppen im Schnitt nicht voneinander – obwohl in beiden Sprachregionen auf Englisch längere Texte geschrieben worden waren (vgl. Abbildung 2).

[11] HD-D ist ab einer Textlänge von 40-50 Tokens ebenfalls ausreichend stark gestreut. In unseren Korpora sind so "lange" Texte aber in erster Linie von Lernenden in ihrer Schulsprache geschrieben wurden. Der Nutzen des Masses für die Analyse kurzer fremdsprachlicher Texte auf niedrigen Kompetenzniveaus ist daher eingeschränkt.

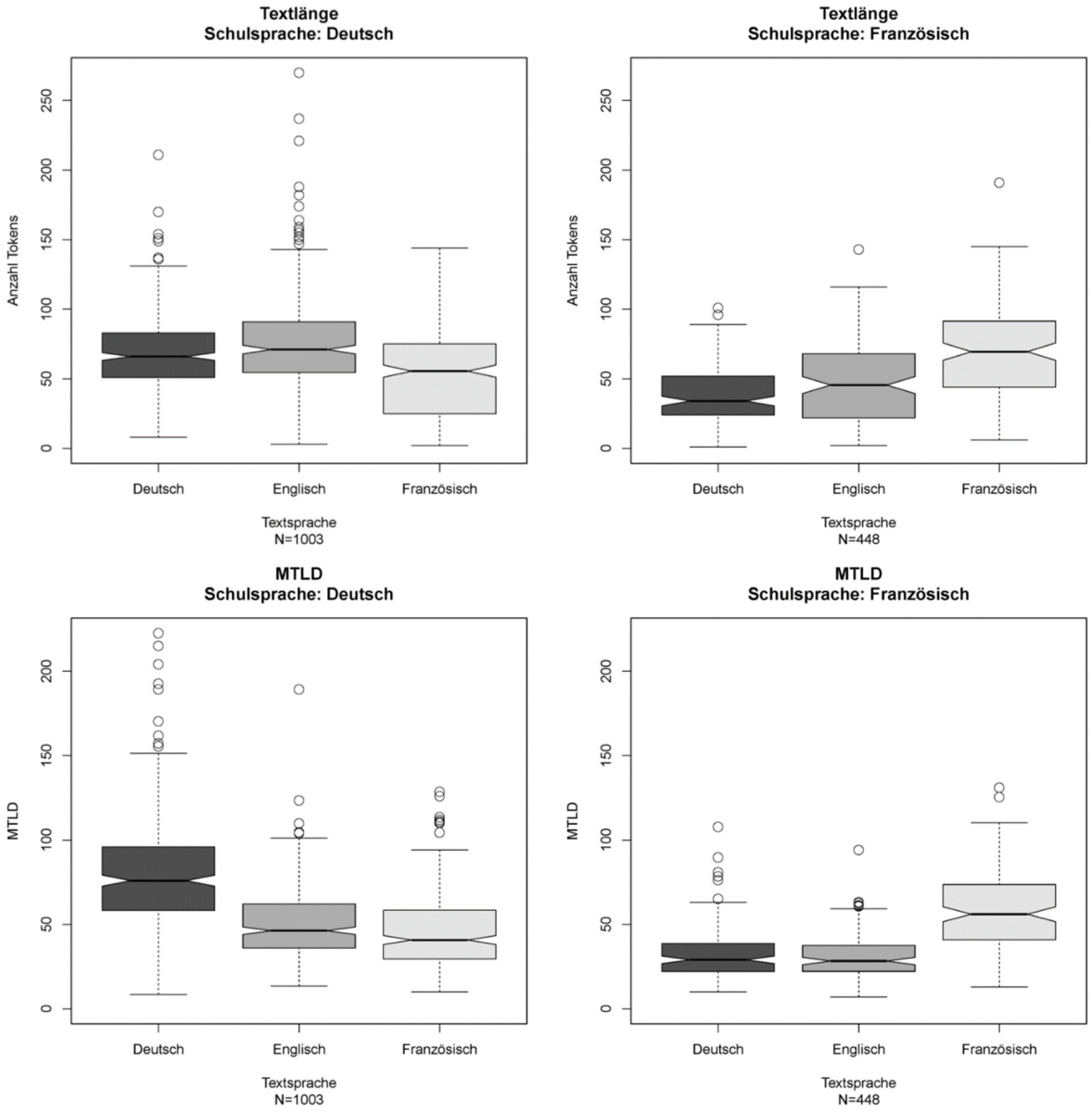

Abbildung 2: Textlänge (oben) und MTLD (unten) pro Sprachenkombination.

Betrachtet man die MTLD-Werte der unterschiedlichen Aufgaben, so zeigt sich ebenfalls, dass diese in den beiden Schulsprachen eine breite Streuung aufweisen, in den beiden Fremdsprachen hingegen deutlich homogener ausfallen.

Schliesslich sei noch eine Beobachtung erwähnt, die aufgrund der Kontrolle der Verteilung entstanden ist: Die MTLD-Werte der deutschsprachigen Texte lassen eine fast eindeutige Zuordnung der Texte zur Gruppe der schulsprachlichen bzw. der fremdsprachigen Texte zu (Abbildung 3). Texte mit einem MTLD-Wert unter 40 wurden fast ausschliesslich von DaF-Lernenden geschrieben, solche zwischen 40 und 60 sind nicht eindeutig zuzuordnen und MTLD-Werte über 60 finden sich fast nur noch in Texten von Lernenden aus

der Deutschschweiz. In den anderen Sprachen ist ein ähnlicher Effekt zu sehen, er ist aber v.a. im Französischen weniger deutlich.

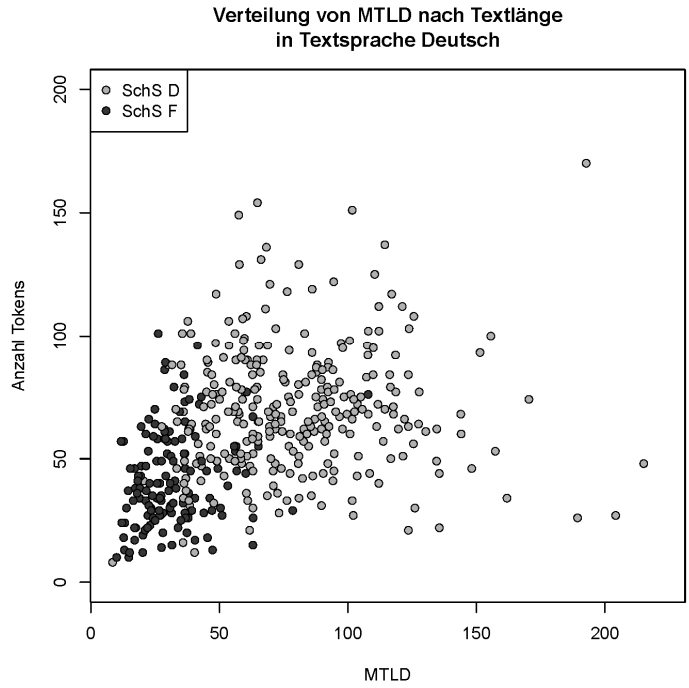

Abbildung 3: Verteilung der Texte in deutscher Sprache nach MTLD und Textlänge.

4.2.2 Korrektheit

Insgesamt lassen sich nur wenige relevante Unterschiede bezüglich der Korrektheit in den Texten finden. Bei der grossen Mehrheit der Texte wurden über 80% aller Tokens orthographisch korrekt geschrieben. Texte in den Schulsprachen sind in beiden Sprachgruppen am korrektesten, gefolgt von Englisch. Die Texte in der jeweiligen ersten Fremdsprache sind insgesamt zu einem grösseren Teil fehlerhaft, insbesondere, wenn man Sprachmaterial in anderen Sprachen hinzunimmt. In diesem Fall enthält nur etwa ein Viertel aller Texte zu mehr als 80% korrekte Tokens.

Eine weitere Beobachtung betrifft die Klassenzugehörigkeit der Lernenden: In der Schulsprache Französisch sind die Texte einer Klasse auffällig korrekter als jene in den anderen drei Klassen. Möglicherweise achtete hier eine Lehrperson stärker auf Rechtschreibung. Bei den sechs Deutschschweizer Klassen zeigt sich wiederum ein klarer Unterschied zwischen den drei Schulen, möglicherweise aufgrund der unterschiedlichen Leistungsniveaus: Insbesondere die Texte aus den gymnasialen Klassen sind in allen drei Sprachen auffällig korrekter als jene der anderen Lernenden.

4.2.3 Wortschatzbreite

Da die verwendeten Referenzkorpora in ihrer Struktur unterschiedlich sind, sind die Masse, die zur Erfassung der Wortschatzbreite erhoben wurden (der Anteil

der Wörter am Text, der zu den häufigsten 1000 Wörtern gehört sowie der durchschnittliche Rang der verwendeten Wörter im Referenzkorpus), nicht geeignet, um Vergleiche zwischen den Sprachen durchzuführen.

Der innersprachliche Vergleich der Wortschatzbreite in den englischen Texten befördert aber ein potenziell interessantes Ergebnis zutage: Es zeigt sich, dass die französischsprachigen Lernenden zwar insgesamt kürzere Texte mit geringeren MTLD-Werten schrieben, dafür aber einen kleineren Anteil an den häufigsten 1000 Wörtern sowie durchschnittlich höhere Rangwerte im Referenzkorpus aufweisen. Einfacher ausgedrückt: Die englischen Texte der französischsprachigen Lernenden weisen leicht (!) selteneren Wortschatz auf als jene der deutschsprachigen Lernenden.

In den anderen beiden Sprachen zeigt sich bezüglich der Wortschatzbreite kein sichtbarer Unterschied, auch nicht zwischen schulsprachlichen und fremdsprachlichen Texten. Dies gilt allerdings bei einer ohnehin grossen Streuung der Werte über alle Texte hinweg.

4.2.4 Wortschatzqualität

Anders als die bisher diskutierten Masse lassen sich die Ergebnisse der Frequenzanalysen nicht als Zahlen ausdrücken. Stattdessen entstanden Listen von Lemmata bzw. Tokens, die in einer Gruppe von Texten (aussergewöhnlich) häufig auftreten. Diese können nur deskriptiv analysiert werden, wobei gerade dieses beschreibende Vorgehen Merkmale der Texte aufdeckt, welche durch die oben beschriebenen Zahlenwerte möglicherweise nicht erfasst werden konnten.

Betrachtet man die am häufigsten auftretenden Tokens und Lemmata in den Texten, die in der jeweiligen Fremdsprache geschrieben wurden, zeigt sich vor allem der Einfluss der Aufgabenstellungen: Vor allem im Französischen und im Deutschen finden sich schon unter den häufigsten 30 Lemmata Inhaltswörter wie *aimer*, *animal*, *vacance(s)*, *mögen*, *Katze*, *heissen* und *Jahr*. Im Vergleich zu den Referenzkorpora treten hingegen bestimmte Funktionswörter weniger auf (z.B. die Pronomen *tu* und *vous* oder die Präpositionen *als*, *zu* und *von*). Im Englischen ist der Effekt weniger ausgeprägt, dort stimmt die Liste der 30 häufigsten Lemmata weitgehend mit jener aus dem COCA-Korpus überein. Vergleicht man weiterhin die 50 häufigsten Lemmata in den englischen Texten der deutschsprachigen Lernenden mit jenen der französischsprachigen Lernenden, so zeigt sich, dass Erstere die folgenden Wörter häufiger nutzten: *time*, *more*, *think*, *should*, *thing*, *or*, *on*, *would* und *also*.

Vergleicht man die Texte in den Fremdsprachen mit den Texten in der gleichen Sprache als Schulsprache, so zeigt sich bezüglich der Keywords ein recht deutlicher Unterschied: Benutzen die DaF-Lernenden vor allem überdurchschnittlich viele Inhaltswörter, die mit den Aufgabenstellungen und vor allem ihrer eigenen Person zu tun haben (*Lieblingshobby*, *Hockey*, *Oma*, *Fussball*),

so finden sich in den schulsprachlichen Texten vor allem Funktionswörter (*welche*, *noch*, *doch*, *was nach*, *man*, *dafür, gar*, *sich*) sowie einzelne Inhaltswörter (*Städtereise*, *Haushalt*, *Erfindung*), die ebenfalls mit den Aufgabenstellungen zusammenhängen, von den Deutschlernenden aber nicht (bzw. seltener) verwendet wurden.

Im Französischen zeigt sich ein ähnliches Bild: Neben eher anspruchsvollen Inhaltswörtern wie *réseau*, *équitation*, *jeunesse* und *dépenser* finden sich in den schulsprachlichen Texten vor allem Funktionswörter wie *où*, *surtout*, *avant*, *ensuite*, *mieux* und besonders auch das Wort *certain*, welches den Französisch-lernenden offensichtlich weniger geläufig ist.

Im Teilkorpus der englischen Texte wurden die Texte der deutschsprachigen Lernenden mit jenen der französischsprachigen verglichen. Wie bereits weiter oben angedeutet, zeigt sich zwischen diesen beiden Gruppen von Lernenden ein Unterschied, der mit jenem zwischen Schul- und Fremdsprache vergleichbar ist: Während die *Token-Keywords* in den Texten aus der Deutschschweiz in erster Linie Funktionswörter und Verben sind, sind die *Keywords* in den französischsprachigen Texten nur einige wenige Inhaltswörter, die in den Aufgaben zentral sind (vgl. Abbildung 4).

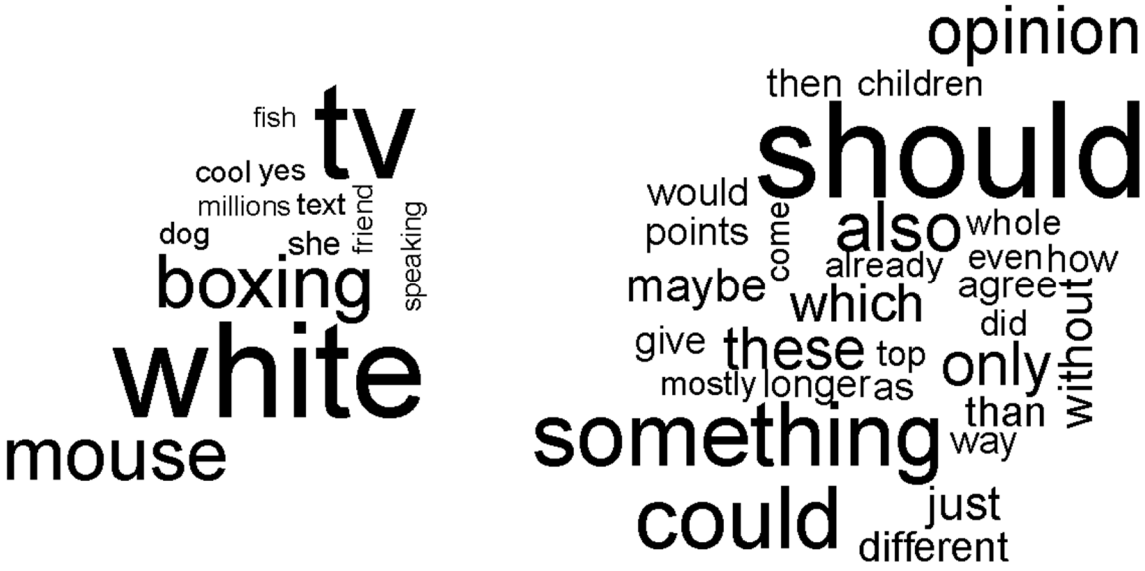

Abbildung 4: Wortwolken der wichtigsten Keywords in den englischsprachigen Texten der beiden Sprachgruppen (Französisch links, Deutsch rechts).

4.3 Crosslinguistic influence

Bei der Verwendung nicht-zielsprachlicher Wörter in den fremdsprachlichen Texten der Schülerinnen und Schüler zeigen sich u.a. Unterschiede in den beiden Sprachregionen: Die von den Deutschschweizer Lernenden auf Französisch verfassten Texte weisen aufgabenübergreifend zwischen 1% und

5% deutsches Wortmaterial auf, englische Wörter finden sich in diesen Texten dagegen nur am Rande (Werte unter 1%). Bei den englischen Texten sind anderssprachige Einflüsse generell sehr selten, wobei die Schulsprache Deutsch nur eine marginale und die erste Fremdsprache Französisch gar keine Rolle spielt (vgl. am Beispiel eines Tasks Tabelle 2).

	Anzahl Tokens	CLI Englisch	CLI Französisch	CLI Schulsprache
Französische Texte	1871	3 (< 1%)		59 (3%)
11H-B (14 SuS)	869	1		48
11H-A (10 SuS)	284	2		7
1. MatS (10 SuS)	718	0		4
Englische Texte	2877		1 (< 1%)	21 (<1%)
11H-B (13 SuS)	965		0	12
11H-A (15 SuS)	873		1	5
1. MatS (11 SuS)	1039		0	4

Tabelle 2: Nicht-zielsprachliche Einflüsse in den fremdsprachlichen Texten der Deutschschweizer Schülerinnen und Schüler am Beispiel der Aufgabe SWI04[12].

Tabelle 2 zeigt auch, dass die Transfers aus der Schulsprache (hier Deutsch), die insgesamt sehr selten sind, zum grossen Teil von den jüngeren Schülerinnen und Schülern im mittleren Leistungsniveau ausgehen.

Tabelle 3 zeigt die Ergebnisse der gleichen Analyse aus der Westschweiz. Darin wird deutlich, dass es in den in Deutsch als Fremdsprache verfassten Texten vergleichsweise viel französisches und etwas englisches Material gibt. Allerdings sind diese Werte als Ausreisser einzustufen: Je nach Aufgabe schwankt der Anteil der Schulsprache zwischen 5 und 10% und ist damit etwa doppelt so hoch wie bei den französischen Texten aus der Deutschschweiz. Auch die Transfers aus dem Englischen sind in den deutschen Texten der französischsprachigen Lernenden zahlreicher als in den französischen Texten der Deutschschweiz. Wie schon in der Deutschschweiz sind die Einflüsse der Schulsprache bei den englischen Texten zwar selten, aber vorhanden, jene aus der ersten Fremdsprache hingegen praktisch inexistent.

[12] Zu den Abkürzungen vgl. Fussnote 6; zusätzlich wird hier und in Tabelle 3 bei den HarmoS-Klassenstufen nach dem höheren (A) bzw. mittleren Leistungsniveau (B) unterschieden.

	Anzahl Tokens	CLI Englisch	CLI Französisch	CLI Schulsprache
Deutsche Texte	559	25 (4.5%)		75 (13%)
10H-B (12 SuS)	380	10		58
10H-B (8 SuS)	179	15		17
Englische Texte	681		5 (<1%)	31 (5%)
10H-B (11 SuS)	435		1	23
10H-B (7 SuS)	246		4	8

Tabelle 3: Nicht-zielsprachliche Einflüsse in den fremdsprachlichen Texten der Westschweizer Schülerinnen und Schüler am Beispiel der Aufgabe SWI04[13].

Bei all diesen Tendenzen ist zu beachten, dass es beträchtliche Variation in den Daten gibt, zwischen individuellen Lernenden, Klassen, Klassenstufen und nicht zuletzt auch zwischen den Aufgaben. So fallen bei allen acht Aufgaben Lernende auf, deren fremdsprachliche Texte keinerlei anderssprachliche Einflüsse erkennen lassen, während bei anderen das Gegenteil zu beobachten ist: Einige Lernende scheinen ihr mehrsprachiges Repertoire recht intensiv auszuschöpfen, dies allerdings nicht immer in allen Aufgaben.

5. Diskussion

5.1 Textlänge und Wortschatz

Die oben beschriebenen Beobachtungen bezüglich der "objektiv" feststellbaren Textmerkmale (Textlänge und verschiedene Wortschatzmasse) lassen zwei zentrale Schlussfolgerungen zu: Zum einen unterscheiden sich schulsprachliche Texte hinsichtlich fast aller Merkmale klar von den fremdsprachlichen Texten in der jeweils gleichen Sprache. Die Texte in der Schulsprache sind länger, korrekter und, auf den Wortschatz bezogen, vielfältiger. Häufige und im Vergleich häufigere Wörter betreffen vor allem Funktionswörter, die der Textstrukturierung und dem präziseren Ausdruck dienen können. Zum anderen sind die Texte der beiden Sprachgruppen in ihrer Fremdsprache Englisch unterschiedlich – die verschiedenen Indikatoren, mit Ausnahme der Wortschatzhäufigkeit (vgl. 4.2.3)[14], weisen darauf hin, dass die deutschsprachigen Lernenden im Englischen insgesamt über eine höhere

[13] Zu den Abkürzungen vgl. Fussnoten 6 und 12.

[14] Dieses anomale Ergebnis hat möglicherweise mit der Zusammensetzung des englischen Wortschatzes zu tun: Wörter, die aus der romanischen Sprachfamilie stammen, sind im englischen tendenziell seltener, für Sprecherinnen und Sprecher romanischer Sprachen aber leichter zu lernen. Es kann also sein, dass die Wortschatzhäufigkeit im Englischen zwischen Lernenden mit Deutsch und Französisch als Schulsprache nicht direkt vergleichbar ist (vgl. auch Meara 1996: 43).

Sprachkompetenz verfügen als die Lernenden aus der Westschweiz. Die Profile der beteiligten Klassen sprechen stark für diese Vermutung: Hatten in der Westschweiz vier Klassen des mittleren Anforderungsprofils der 2. Sekundarklasse (10H) teilgenommen, waren die Lernenden in der Deutschschweiz bereits in der 3. Sekundarklasse (11H) und zumindest teilweise im höchsten Anforderungsprofil. Dort nahm sogar eine 1. Klasse einer Maturitätsschule teil. Die Autorinnen und Autoren der Texte von 2018 waren also älter, besuchten z. T. anspruchsvolleren Unterricht und hatten, dies verrät ein Blick auf die Stundentafeln der Kantone, mehr Englischunterricht.

Legt man diese Feststellung zugrunde, so ergibt sich auch ein Erklärungsansatz für die unterschiedlichen Beobachtungen zur Textlänge in den beiden Sprachregionen (vgl. Abschnitt 4.1): Einerseits ist die auffällig ähnliche Länge aller Texte aus der Deutschschweiz wohl schlicht durch die Vorgabe einer Wortanzahl in den Aufgabenstellungen zu begründen: In sämtlichen Aufgaben wurden die Lernenden aufgefordert, 60-80 Wörter zu schreiben. Wir vermuten, dass die Lernenden sich nach diesen Vorgaben richteten, sofern sie dafür genügend sprachliches Material zur Verfügung hatten[15].

Auf ähnliche Weise liesse sich wohl auch die jeweils unterschiedliche Textlänge der am Computer und auf Papier geschriebenen Texte in den beiden Sprachregionen begründen: Falls die Lernenden in der Westschweiz in ihren beiden Fremdsprachen tatsächlich eher geringere Kompetenzen hatten, die kaum für die Erfüllung der Wortvorgabe genügten, so könnte sich dies erschwerend auf das Schreiben am Computer ausgewirkt haben. Sobald die fremdsprachliche Schreibkompetenz der Lernenden jedoch genügte, was in der Deutschschweiz mehrheitlich der Fall zu sein scheint, könnte der durch das Medium erzeugte Unterschied verschwunden sein.

Um diese Vermutungen zu prüfen, wird es aber notwendig sein, die Texte systematisch zu raten und die Analysen durch detailliertere Fehlerannotationen zu ergänzen.

5.2 Crosslinguistic influence

Die berichteten Befunde zur Verwendung verschiedener Sprachen in den fremdsprachlichen Schülertexten vermitteln einen ersten Eindruck davon, in welchem Ausmass die Lernenden ihr sprachliches Repertoire beim Schreiben einbringen. Als Tendenz lässt sich festhalten, dass die älteren Lernenden in der Deutschschweiz in den französischen Texten wenig auf ihre Schulsprache und das Englische zurückgreifen und in den englischen Texten noch weniger anderssprachliches Material nutzen. Dagegen greifen die jüngeren Schüler-innen und Schüler aus der Westschweiz beim Schreiben auf Deutsch recht

[15] Eine ähnliche Beobachtung machte auch Chambers (2008: 11), die Textlängen um 100 Wörter sowohl auf Papier als auch am Computer fand, nachdem die Testteilnehmenden aufgefordert worden waren, Texte von etwa 100 Wörtern zu schreiben.

häufig auf ihre Schulsprache und seltener auch auf Englisch zurück. Auch beim Verfassen englischer Texte scheinen die Einflüsse der Schulsprache grösser zu sein als diejenigen des Deutschen. Diese Tendenzen lassen sich kaum mit anderen Studien vergleichen[16], sie werfen aber ein Licht auf die in Abschnitt 2.3 skizzierte Diskussion, ob der oft beobachtete Einfluss einer nahverwandten L2 auf die L3 primär eine Folge typologischer Ähnlichkeit ist oder mehr mit dem L2-Status-Faktor zu tun hat. So könnte man erwarten, dass Deutsch beim Schreiben englischer Texte für die französischsprachigen Lernenden eine im doppelten Sinne geeignete Stützsprache ist: Es ist mit Englisch nahverwandt *und* sollte den Vorteil des L2-Status-Faktors auf seiner Seite haben. Die vorliegenden Daten scheinen dieser Erwartung zu widersprechen: Wenn die Westschweizer Schülerinnen und Schüler beim Schreiben in englischer Sprache auf eine andere Sprache zurückgreifen, ist es primär Französisch, nicht Deutsch. Ebenso wenig scheinen die beiden theoretischen Vorteile bei dieser Zielgruppe beim Schreiben auf Deutsch eine Rolle zu spielen: Auch hier ist der Einfluss von Englisch als nahverwandter Sprache marginal und jedenfalls kleiner als der Einfluss der Schulsprache. Das verweist darauf, dass im gegebenen Kontext Faktoren wie Sprachunterricht und Kontakt zu den beteiligten Sprachen, wahrgenommene Sprachähnlichkeit und, ganz besonders, Sprachkompetenzen im Sinne der ausreichenden Transferbasis wichtiger sein könnten als die (formal-)typologische Ähnlichkeit und der L2-Status-Faktor.

6. Schluss

Die grundlegende Frage des vorliegenden Artikels war, inwiefern mittels quantitativer Methoden erfassbare Textmerkmale Auskunft geben über die Sprachkompetenz der Lernenden. Es zeigte sich, dass die verschiedenen Indikatoren, angefangen bei der Textlänge, unter bestimmten Bedingungen und über viele Texte hinweg eine erste Kategorisierung durchaus erlauben. So scheint die Textlänge, sofern diese nicht durch die Aufgabenstellung gedeckelt ist, grob zwischen qualitativ besseren und weniger guten Texten zu unterscheiden. Dafür spricht, dass die meisten der anderen untersuchten Indikatoren in eine ähnliche Richtung zeigen. So haben v.a. schulsprachliche Texte auch höhere MTLD-Werte und niedrigere Fehlerquotienten. Sofern die Masse zuverlässig sind, würde dies bedeuten, dass längere Texte tendenziell mit höherer Wortschatzvielfalt und weniger Fehlern geschrieben werden und daher auf eine höhere Schreibkompetenz hinweisen. Ein qualitativer Blick auf die in den Texten verwendeten Wörter erhärtet diesen Eindruck. Auch der

[16] Christen & Näf (2001) analysierten 400 Deutschtexte von frankophonen Lernenden aus dem Korpus der DiGS-Studie (Diehl et al. 2000) und fanden 376 Transfers aus dem Englischen. Da aber nicht ganz deutlich wird, wie lang die analysierten Texte waren, lassen sich diese Befunde mit den vorliegenden nicht genau vergleichen. Deckungsgleich ist hingegen die Feststellung sehr grosser individueller Variation in den Schülertexten.

zunehmend geringere Anteil nicht-zielsprachlicher Wörter spricht in unseren Daten eher für höhere Sprachkompetenz.

Gleichzeitig zeigen die Resultate klar, dass die Streuung der erhobenen Masse meist sehr hoch ist und auch bei CLI grosse individuelle Unterschiede zu beobachten sind. Es ist daher zu vermuten, dass aus einem erhobenen Wert kaum auf die Qualität von einzelnen Texten geschlossen werden kann. Genauere Hinweise könnte hier eine statistische Analyse liefern, die Werte wie Textlänge, MTLD und Indikatoren für Wortschatzbreite auch auf der Ebene der einzelnen Texte und Lernenden berücksichtigt sowie Hintergrundvariablen wie die Zugehörigkeit zu einer Klasse miteinbezieht. Ebenso bleibt ein menschliches Rating ein Desiderat, da dieses weitere Informationen zur Qualität einzelner Texte liefern kann.

Schliesslich zeigen die Daten aus den beiden Sprachregionen auch, dass die Auswirkungen von Aufgaben und Aufgabenstellungen sowie Erhebungsbedingungen stark davon abhängig sind, über welche Sprachkompetenzen die Lernenden verfügen, in welchen Gruppen sie zusammen lernen und welche Lernkulturen ihnen vertraut sind. Anders als in einer früheren Analyse finden sich nun, wohl aufgrund der stärkeren Heterogenität der Lernenden, kaum noch verallgemeinerbare Gruppenunterschiede, z.B. zwischen Texten, die am Computer und auf Papier geschrieben wurden. Dieser Befund unterstreicht die Notwendigkeit, Schreibkompetenz als ein komplexes System von interdependenten Variablen zu sehen, für das es zur Zeit kein einheitliches Modell zu geben scheint, sodass das System immer nur in Teilen und unter bestimmten Einschränkungen abgebildet und gemessen werden kann. Dies gilt für Schlussfolgerungen grosser korpuslinguistischer Studien ebenso wie für schriftliche Prüfungen im Schulkontext.

Die Autoren danken Katia Rey, zwei anonymen Reviewern und dem Redaktionsteam der Sondernummer für ihre wertvollen Hinweise und Kommentare zu früheren Versionen dieses Artikels.

LITERATUR

Bachman, L. F. & Palmer, A. S. (2010). *Language assessment in practice. Developing language assessments and justifying their use in the real world*. Oxford, New York: Oxford University Press.

Bärenfänger, O., Harsch, C., Tesch, B. & Vogt, K. (2019). Reform, Remake, Retusche? – Diskussionspapier der Deutschen Gesellschaft für Fremdsprachenforschung zum Companion to the CEFR (2018). *Zeitschrift für Fremdsprachenforschung, 30(1)*, 7-13.

Becker-Mrotzek, M. & Schindler, K. (2007). *Schreibkompetenz modellieren*. In M. Becker-Mrotzek & K. Schindler (Hgg.), *Texte schreiben* (S. 7-26). Duisburg: Gilles & Francke.

Bleyhl, W. (1996). Fremdsprachenfachdidaktik/Sprachlehr- und -lernforschung als wissenschaftliche Disziplin. Zwischenbilanz 1996. In K.-R. Bausch, H. Christ, F. G. Königs & H.-J. Krumm (Hgg.), *Erforschung des Lehrens und Lernens fremder Sprachen: Zwischenbilanz und Perspektiven. Arbeitspapiere der 16. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts* (S. 19-30). Tübingen: Gunter Narr.

Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, I(1)*, 1-47.

Chambers, L. (2008). Computer-based and paper-based writing assessment: A comparative text analysis. *Cambridge ESOL Research Notes, 34*, 9-15.

Chen, X. & Meurers, D. (2016). CTAP: A web-based tool supporting automatic complexity analysis. In *Proceedings of the workshop on computational linguistics for linguistic complexity (CL4LC)* (S. 113-119). Osaka: The COLING 2016 Organizing Committee.

Christen, H. & Näf, A. (2001). Trausers, shoues und Eis. Englisches im Deutsch von Französischsprachigen. In K. Adamzik & H. Christen (Hgg.), *Sprachkontakt, Sprachvergleich, Sprachvariation* (S. 61-98). Tübingen: Niemeyer.

Council of Europe (2018). *Common European framework of reference for languages: learning, teaching, assessment. Companion volume with new descriptors*. Strasbourg: Council of Europe.

Diehl, E., Christen, H., Studer, T., Leuenberger, S. & Pelvat, I. (2000). *Grammatikunterricht: Alles für der Katz?* Tübingen: Niemeyer.

Eberharter, K., Kremmel, B. & Konzett-Firth, C. (2018). Produktive Fertigkeiten überprüfen und bewerten. In B. Hinger & W. Stadler (Hgg.), *Testen und Bewerten fremdsprachlicher Kompetenzen. Eine Einführung* (S. 87-115). Tübingen: Narr.

Ecke, P. (2015). Parasitic vocabulary acquisition, cross-linguistic influence, and lexical retrieval in multilinguals. *Bilingualism: Language and Cognition, 18(2)*, 145-162.

Efeoglu, G., Yüksel, H. G. & Baran, S. (2019). Lexical cross-linguistic influence: a study of three multilingual learners of L3 English. *International Journal of Multilingualism* 1-17.

Europarat (2001). *Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, lehren, beurteilen*. Berlin: Langenscheidt.

Grabowski, J., Blabusch, C. & Lorenz, T. (2007). Welche Schreibkompetenz? Handschrift und Tastatur in der Hauptschule. In M. Becker-Mrotzek & K. Schindler (Hgg.), *Texte schreiben* (S. 41-61). Duisburg: Gilles & Francke.

Harsch, C., Neumann, A., Lehmann, R. & Schröder, K. (2007). Schreibfähigkeit. In E. Klieme & B. Beck (Hgg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)* (S. 42-62). Weinheim, Basel: Beltz.

Hayes, J. R. & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Hgg.), *Cognitive Processes in Writing* (S. 3-30). Hillsdale: Lawrence Erlbaum.

Housen, A. & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30(4)*, 461-473.

Jarvis, S. (2013). Defining and measuring lexical diversity. In M. Daller & S. Jarvis (Hgg.), *Vocabulary knowledge: Human ratings and automated measures*. Amsterdam, Philadelphia: John Benjamins.

Jessner, U. (2008). Teaching third languages: Findings, trends and challenges. *Language Teaching, 41(1)*, 15-56.

Karges, K., Barras, M. & Lenz, P. (i. V.). Assessing young language learners' receptive skills: Should we ask the questions in the language of schooling? In S. Frisch, E. Romeik & J. Rymarczyk (Hgg.), *Current research into young foreign language and English as an additional language learners' literacy skills*. Berlin: Peter Lang.

Karges, K., Studer, T. & Wiedenkeller, E. (2019). On the way to a new multilingual learner corpus of foreign language learning in school: observations about task variation. In A. Abel, A. Glaznieks, V. Lyding & L. Nicolas (Hgg.), *Widening the scope of learner corpus research. Selected papers from the fourth Learner Corpus Research Conference* (S. 137-165). Louvain-la-Neuve: Presses universitaires de Louvain.

Kellogg, R. T., Whiteford, A. P., Turner, C. E., Cahill, M. & Mertens, A. (2013). Working memory in written composition: An evaluation of the 1996 model. *Journal of Writing Research, 5(2)*, 159-190.

Kim, M., Crossley, A. & Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *The Modern Language Journal, 102(1)*, 120-141.

Konsortium HarmoS Fremdsprachen (2009). *Fremdsprachen. Wissenschaftlicher Kurzbericht und Kompetenzmodell (provisorische Fassung)*. http://www.edudoc.ch/static/web/arbeiten/harmos/L2_wissB_25_1_10_d.pdf. Zuletzt abgerufen am 18.02.2020.

Kyle, K. & Crossley, A. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing, 34*, 12-24.

Levelt, W. J. M. (1993). *Speaking. From intention to articulation*. Cambridge, London: The MIT Press.

McCarthy, P. M. & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42(2)*, 381-392.

McNamara, D. S., Graesser, A. C., McCarthy, P. M. & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press.

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer & J. Williams (Hgg.), *Performance and competence in second language acquisition* (S. 35-53). Cambridge: Cambridge University Press.

Meisel, J. (1983). Transfer as a second language strategy. *Language and Communication, 3*, 11-46.

Michalke, M. (2017). *koRpus: An R package for text analysis*. Düsseldorf: reaktanz.de.

Michalke, M. (2019). *Package "koRpus"*. https://reaktanz.de/R/pckg/koRpus/koRpus.pdf. Zuletzt abgerufen am 18.02.2020.

Odlin, T. (2012). Crosslinguistic influence in Second Language Acquisition. In C. A. Chapelle (Hg.), *The Encyclopedia of Applied Linguistics*. Hoboken: Wiley.

Olinghouse, N. G. & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing, 26(1)*, 45-65.

Petersen, I. (2019). Messung, Beurteilung und Förderung von Schreibkompetenz in Deutsch als Erst- und Zweitsprache – ein Überblick. In I. Kaplan & I. Petersen (Hgg.), *Schreibkompetenzen messen, beurteilen und fördern* (S. 11-36). Münster: Waxmann.

Polio, C. & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing, 26*, 10-27.

R Core Team (2017). *R: A language and environment for statistical computing*. Wien: R Foundation for Statistical Computing.

Schmid, H. (2018). *TreeTagger - a language independent part-of-speech tagger*. http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/. Abgerufen am 21.02.2018.

Schneider, G. (2007). Fremdsprachenforschung und die Ausbildung von Fremdsprachenlehrerinnen und -lehrern. *Beiträge zur Lehrerinnen- und Lehrerbildung, 25(2)*, 143-155.

Singleton, D. (2012). Multilingual lexical operations. In J. Cabrelli Amaro, S. Flynn & J. Rothman (Hgg.), *Third language acquisition in adulthood* (S. 95-113). Amsterdam: John Benjamins.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30(4)*, 510-532.

Slabakova, R. (2017). The scalpel model of third language acquisition. *International Journal of Bilingualism, 21(6)*, 651-665.

Studer, T. (i. V.). Jetzt skaliert! Plurikulturelle und mehrsprachige Kompetenzen im erweiterten Referenzrahmen. *Deutsch als Fremdsprache 2020(1)*.

Treffers-Daller, J. (2009). Code-switching and transfer: an exploration of similarities and differences. In B. E. Bullock & A. J. Toribio (Hgg.), *The Cambridge handbook of linguistic code-switching* (S. 58-74). Cambridge: Cambridge University Press.

Treffers-Daller, J. (2018). The measurement of bilingual abilities: central challenges. In A. De Houwer & L. Ortega (Hgg.), *The Cambridge handbook of bilingualism* (S. 289-306). Cambridge: Cambridge University Press.

Treffers-Daller, J., Parslow, P. & Williams, S. (2018). Back to basics: How measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics, 39(3)*, 302-327.

Vasylets, O., Gilabert, R. & Manchón, R. M. (2017). The effects of mode and task complexity on second language production. *Language Learning, 67(2)*, 394-430.

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Hgg.), *Definition and selection of competencies. Theoretical and conceptual foundations* (S. 33-66). Göttingen: Hogrefe & Huber Publishers.

Weir, C. J. & Khalifa, H. (2008). A cognitive processing approach towards defining reading comprehension. *Cambridge ESOL Research Notes, 31*, 2-10.