

Universität Freiburg (CH)

Departement für Mehrsprachigkeitsforschung und Fremdsprachendidaktik

Masterarbeit

Eingereicht bei der Philosophischen Fakultät der Universität Freiburg (CH)

Lexikalische Vielfalt in schriftlichen Produktionen von Schülerinnen und Schülern

**Eine vergleichende Untersuchung von subjektiver Einschätzung und
ausgewählten Massen lexikalischer Vielfalt in den Fremdsprachen Französisch
und Englisch am Ende der obligatorischen Schulzeit**

Arbeit eingereicht bei: Prof. Dr. Raphael Berthele

Abgabedatum: 23.06.2023

Sebastian Stephan Salzmann

Heimatort: Naters

10-933-984

Leischenstrasse 25

3945 Gampel

sebastian.salzmann@unifr.ch

Inhaltsverzeichnis

1 Einleitung	4
2 Theoretischer Rahmen	7
2.1 <i>Sprachlicher Bildungsstandard in der obligatorischen Schule</i>	7
2.1.1 Gemeinsamer Europäischer Referenzrahmen für Sprachen (GER)	8
2.2 <i>Lehrplan 21 für Fremdsprachen in der Deutschschweiz</i>	9
2.3 <i>Schreibkompetenzen in der ersten und zweiten Fremdsprache im Vergleich</i>	11
2.4 <i>Lexikalischer Reichtum – Lexikalische Vielfalt</i>	12
2.4.1 Terminologie.....	13
2.4.2 Operationalisierung der lexikalischen Vielfalt	14
3 Empirischer Teil.....	24
3.1 <i>Forschungsfrage</i>	24
3.2 <i>Hypothesen.....</i>	24
3.3 <i>Design</i>	28
3.4 <i>Methode</i>	29
3.4.1 Textsammlung	29
3.4.2 Ratings	38
3.5 <i>Ergebnisse.....</i>	45
4 Diskussion	62
4.1 <i>Hypothese 1</i>	62
4.2 <i>Hypothese 2</i>	63
4.3 <i>Hypothese 3</i>	65
4.4 <i>Hypothese 4</i>	65
4.5 <i>Hypothese 5</i>	66
4.6 <i>Hypothese 6</i>	69
4.7 <i>Hypothese 7</i>	70
5 Fazit.....	71
6 Literaturverzeichnis.....	77
7a Abbildungsverzeichnis	85

7b Tabellenverzeichnis	85
8 Anhang.....	86
<i>8.1 Aufgabenstellung Schülerinnen und Schüler</i>	<i>86</i>
8.1.1 Pilotphase zur Erhebung der Texte.....	86
8.1.2 Beispiel Aufgabenstellung zur Erhebung der Texte in Englisch	87
8.2 Fragebogen Schülerinnen und Schüler	88
8.3 Übersicht Lehrperson.....	94
8.5. Videoclips.....	95
8.6 Berechnungen Abbildung 1	96
8.7 Berechnungen Abbildung 2	96
8.8 Berechnungen Abbildung 8	96
8.9 Berechnungen Abbildung 9	97
8.10 Lexikalische Vielfalt: Guiraud/MTLD vs. Rating	97
9 Eidesstattliche Erklärung	98

1 Einleitung

Die vorliegende Arbeit beschäftigt sich mit dem Fremdsprachenunterricht im deutschsprachigen Teil des Kantons Wallis und untersucht anhand verschiedener Masse der lexikalischen Vielfalt Textproduktionen von Schülerinnen und Schülern in der Unterrichtssprache und der ersten Fremdsprache Französisch, sowie der zweiten Fremdsprache Englisch. Der Kontext dieser Untersuchung ist ein für diese Untersuchung generierter Textkorpus, der von zukünftigen Lehrpersonen beurteilt wurde. Dabei wurden verschiedene Merkmale des Textes bewertet, darunter die lexikalische Vielfalt, die Erfüllung der Aufgabenstellung, die Grammatik und die Rechtschreibung.

Die Schweizer Bildungslandschaft hat sich in den letzten Jahren sprachpolitisch stark verändert. In der interkantonalen Vereinbarung über die Harmonisierung der obligatorischen Schule, dem HarmoS-Konkordat von 2008 (HarmoS-Konkordat, 2008, Art. 4), wurde der Fremdsprachenunterricht geregelt. Zudem hat die Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK) bereits 2004 beschlossen (vgl. EDK, 2004), dass alle Lernenden in der obligatorischen Schulzeit zwei Fremdsprachen erlernen sollen, wobei die Reihenfolge der Einführung kantonal geregelt ist. Die erste Fremdsprache wird ab der fünften Klasse (5H) und die zweite ab der siebten Klasse (7H) unterrichtet. Im deutschsprachigen Wallis wird Französisch als erste Fremdsprache und Englisch als zweite Fremdsprache unterrichtet.

Bisherige Untersuchungen im 8. und 10. Schuljahr nach HarmoS zeigen, dass es in der Bildungsregion Zentralschweiz, in der Englisch als erste Fremdsprache und Französisch als zweite Fremdsprache eingeführt wird, deutliche Unterschiede zwischen den Sprachkompetenzen in der ersten und zweiten Fremdsprache gibt, insbesondere im schriftlichen Bereich (Peyer et al., 2016b, S. 53). Laut Studienergebnissen erreicht eine Minderheit der Schülerinnen und Schüler des 10. Schuljahrs das erforderliche Niveau A2.1 in Französisch gemäss dem Gemeinsamen Europäischen Referenzrahmen für Sprachen *GER* (vgl. Europarat, 2001) und dem Lehrplan 21 (vgl. D-EDK, 2016). Rund zwei Drittel der Lernenden im Englischunterricht hingegen erreichen die Lehrplanziele und somit das Niveau A 2.1 nach GER (Peyer et al., 2016a, S. 53). Die vorliegende wissenschaftliche Arbeit hat den Fokus auf den Bereich Schreiben gelegt, speziell in Bezug auf die lexikalische Vielfalt.

Verschiedene Forschende gehen davon aus, dass die Verwendung einer angemessenen Wortwahl einen grossen Einfluss auf die Textqualität hat (z.B. S. A. Crossley & McNamara, 2011; Olinghouse & Wilson, 2013) und suchen daher nach einer Möglichkeit, den in Texten verwendeten Wortschatz objektiv zu bewerten. Wie Jarvis (2013a) feststellt, bleibt jedoch trotz des verbreiteten Einsatzes von Metriken für den lexikalischen Reichtum in der quantitativen Linguistik unklar, wie gut sie Aspekte des lexikalischen Reichtums messen. Daher werden Masse der lexikalischen Vielfalt oft anhand ihrer Sensibilität für die Länge eines Textes (z.B. McCarthy & Jarvis, 2007; Treffers-Daller, 2013; Tweedie & Baayen, 1998) bewertet und in Korrelation mit anderen Variablen, wie beispielsweise der Qualität eines Textes oder der Sprachkompetenz, gesetzt (z.B. H. Daller et al., 2003; Treffers-Daller, 2013; Treffers-Daller et al., 2016).

Jarvis (2013a) schlägt vor, den lexikalischen Reichtum als erkenntnistheoretisches Phänomen zu betrachten und argumentiert, dass Texte von unqualifizierten Gutachterinnen und Gutachtern anhand von Reichtumsmetriken beurteilt werden sollten. Darüber hinaus legt Jarvis (2017, 2013a, 2013b) dar, dass das Phänomen des perzeptiven lexikalischen Reichtums in einem begrenzten Ausmass von verschiedenen Dimensionen (vgl. Kapitel 2.4.2.8) erfasst werden kann.

Für die vorliegende Untersuchung bei Schulabgängerinnen und Schulabgängern wurden Texte von 183 Schülerinnen und Schülern gesammelt. Anschliessend wurden verschiedene Indizes berechnet, die mit der lexikalischen Vielfalt in Verbindung stehen. Diese Indizes wurden dann den Bewertungen von angehenden Lehrpersonen im Zyklus 2 in Bezug auf verschiedene Aspekte gegenübergestellt. Das Hauptziel der Arbeit ist es, Unterschiede in der lexikalischen Vielfalt zwischen den Texten der ersten und zweiten unterrichteten Fremdsprache zu untersuchen. Die Ergebnisse deuten darauf hin, dass erstens die Wahrnehmung des lexikalischen Reichtums in erheblichem Masse mit Hilfe textbasierter Indizes vorhergesagt werden kann. Zweitens zeigen die Ergebnisse Unterschiede in der lexikalischen Vielfalt, auch in Zusammenhang mit anderen Variablen. Dazu gehören weiteren Kriterien des Textratings, die Notendurchschnitte im Bereich Schreiben, die ermittelten Masse lexikalischer Vielfalt eines Referenzkorpus und die Gegenüberstellung zur Unterrichtssprache.

Im ersten Teil der Arbeit werden die theoretischen Grundlagen thematisiert. Dabei spielen die sprachpolitischen Entwicklungen und die Erwartungen des im Untersuchungsgebiet

eingeführten Lehrplans 21, sowie verschiedene Masse, welche zur Erhebung der lexikalischen Vielfalt verwendet werden, eine wichtige Rolle. Zusätzlich wird im ersten Teil dieser Arbeit der Fremdsprachenunterricht im deutschsprachigen Wallis behandelt. Im zweiten Teil der Arbeit wird unter anderem die verwendete Methode beschrieben. Es erfolgt eine Analyse der Textproduktionen hinsichtlich ausgewählter Masse der lexikalischen Vielfalt. Diese Analyse wird für die Unterrichtssprache Deutsch sowie die erste Fremdsprache Französisch und die zweite Fremdsprache Englisch durchgeführt. Dabei wird die deskriptive Statistik verwendet, um die Ergebnisse zu veranschaulichen. Zusätzlich werden Korrelationsdiagramme und lineare Regressionsberechnungen zur weiteren Analyse genutzt. Hierbei werden durch die Gegenüberstellung mit einem Referenzkorpus in der jeweiligen Sprache auch Unterschiede mit Lernenden, die in der jeweiligen Sprache als Unterrichtssprache unterrichtet werden, sichtbar gemacht. Zusätzlich fliessen andere Grössen, wie die durch Fragebögen ermittelten Notendurchschnitte im Bereich Schreiben sowie weitere Aspekte, die beim Rating berücksichtigt werden, in die Auswertung der Daten ein. Im letzten Teil der Arbeit folgt eine Zusammenfassung zentraler Ergebnisse der Untersuchungen, welche anschliessend anhand der definierten Hypothesen diskutiert werden. Das Fazit bildet den Schluss dieser Arbeit.

2 Theoretischer Rahmen

In diesem Kapitel geht es zum einen um die Verortung des Bereichs Schreiben im schulischen Umfeld, insbesondere in Bezug auf sprachenpolitische Konzepte im schweizerischen Kontext und zum anderen um die allgemeinen theoretischen und praktischen Grundlagen, die bei der vorliegenden Untersuchung berücksichtigt werden. Ergänzend dazu werden für die Untersuchung relevante Begriffe definiert und Masse zur Erhebung des lexikalischen Reichtums anhand aktueller Forschungsergebnisse diskutiert.

2.1 Sprachlicher Bildungsstandard in der obligatorischen Schule

In der interkantonalen Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat, 2008) ist in Artikel 4 u. a. der Sprachenunterricht geregelt. In der mehrsprachigen Schweiz ist es wichtig, dass alle Jugendlichen in einer zweiten Landessprache, die auch kulturelle Aspekte beinhaltet, und in Englisch kommunizieren können. Daher hat die Konferenz der kantonalen Erziehungsdirektorinnen und -direktoren *EDK* (vgl. *EDK*, 2004) im Jahr 2004 mit der Sprachenstrategie beschlossen, dass alle Schülerinnen und Schüler zwei Fremdsprachen lernen sollen. Die Schulen unterrichten spätestens ab dem 5. Schuljahr (5H) die erste Fremdsprache und spätestens ab dem 7. Schuljahr (7H) die zweite Fremdsprache, wobei die Reihenfolge der unterrichteten Fremdsprachen regional koordiniert wird und die Qualität und die Entwicklung des Fremdsprachenunterrichts in einer von der *EDK* genehmigten Gesamtstrategie festgelegt sind (vgl. *EDK*, 2004; HarmoS-Konkordat, 2008, S. 3).

In der deutschsprachigen Region des Kantons Wallis, in welcher die vorliegende Untersuchung durchgeführt wird, wird Französisch als erste Fremdsprache ab der Stufe 5H unterrichtet während Englisch als zweite Fremdsprache ab der Klasse 7H gelehrt wird (Departement für Volkswirtschaft und Bildung des Kantons Wallis, 2018, S. 53; *EDK*, 2022, S. 3). Obwohl die Stundentafel mehr Lektionen für die erste Fremdsprache als für die zweite vorsieht (Staatskanzlei Kanton Wallis, 2020, S. 2; vgl. Staatsrat Kanton Wallis, 2014), werden am Ende der obligatorischen Schulzeit im Sinne der Mehrsprachigkeitsdidaktik vergleichbare Kompetenzen in beiden Fremdsprachen erwartet (Bertschy et al., 2018, S. 3; Departement für Volkswirtschaft und Bildung des Kantons Wallis, 2018, S. 63f.; *EDK*, 2011, S. 6f; HarmoS-Konkordat, 2008, S. 3). Es wird davon ausgegangen, dass der schnellere Lernfortschritt in der zweiten Fremdsprache durch die bereits erworbenen Kompetenzen in der ersten

Fremdsprache ermöglicht wird (Departement für Volkswirtschaft und Bildung des Kantons Wallis, 2018, S. 56; EDK, 2011, S. 6).

Im Jahr 2011 wurden die nationalen Bildungsstandards für Fremdsprachen veröffentlicht (EDK, 2011). Auch die neuen sprachregionalen Lehrpläne, wie der Lehrplan 21, wurden in Übereinstimmung mit der Sprachenstrategie entwickelt und spiegeln diese in vielen Aspekten wider. Der Gemeinsame Europäische Referenzrahmen für Sprachen (GER) soll dabei als Bezugspunkt für den Fremdsprachenunterricht dienen.

2.1.1 Gemeinsamer Europäischer Referenzrahmen für Sprachen (GER)

Der Gemeinsame Europäische Referenzrahmen für Sprachen (GER) wurde zwischen 1991 und 2001 vom Europarat entwickelt und dient als Werkzeug für Fachleute im Bildungsbereich. Er soll die Erstellung von Lehrplänen, Lehrmaterialien und Sprachprüfungen unterstützen (Europarat, 2001).

Im Jahr 2009 legte ein wissenschaftliches Konsortium im Auftrag der EDK einen provisorischen Bericht vor, in dem es Bildungsstandards für Fremdsprachenfächer in der obligatorischen Schule ausgearbeitet hat (Lenz, Peter et al., 2009). Diese Standards beziehen sich auf das Ende der Primar- und Sekundarstufe I. Laut den Ergebnissen des Konsortiums werden die vorgeschlagenen Basisstandards im Bereich Schreiben, auf welche sich die vorliegende Arbeit konzentriert, von nur einer geringen Anzahl der Schülerinnen und Schüler erreicht, obwohl die Standards für das Schreiben in der ersten Fremdsprache in der Klasse 8H und in der Klasse 11H generell tiefer angesetzt sind als für die übrigen kommunikativen Sprachfertigkeiten. Die Expertenvorschläge für Basisstandards im Schreibbereich sind anspruchsvoll, und es ist unwahrscheinlich, dass sie von einer grossen Zahl von Lernenden erreicht werden können, es sei denn, es werden bessere Ergebnisse durch neue Unterrichtsmethoden erzielt (Lenz, Peter et al., 2009, S. 54ff.).

Die am Ende des 11. Schuljahres zu erreichenden Grundkompetenzen (EDK, 2011, S. 27) orientieren sich am GER und werden im Bereich Schreiben zur Erreichung des Niveaus A2.1 folgendermassen definiert:

«Die Schülerinnen und Schüler können in kürzeren einfachen Texten über Menschen und vertraute Erfahrungsbereiche schreiben.

Sie können kurz und einfach über persönlich relevante Ereignisse und Erfahrungen berichten und können ganz einfache Geschichten erzählen.

Sie können kurze, einfache Notizen und Mitteilungen schreiben, um persönliche Bedürfnisse zu kommunizieren, und können ganz einfache persönliche Briefe schreiben.»

2.2 Lehrplan 21 für Fremdsprachen in der Deutschschweiz

In der Deutschschweiz wurden bereits vor dem Start des Lehrplans 21 neue Fremdsprachenlehrpläne entwickelt, um die EDK-Sprachenstrategie umzusetzen. Diese Lehrpläne orientierten sich an Kompetenzen und waren in der Zentral- und Ostschweiz vor der Einführung des Lehrplans 21 in Gebrauch. In den zweisprachigen Kantonen und an der Sprachgrenze wurden sie im Rahmen des Projekts "Passepartout" (Bertschy et al., 2018) ab 2006 eingeführt und inzwischen auch vom Lehrplan 21 übernommen. Die für diese Arbeit erhobenen Daten stammen aus dem Kanton Wallis, in welchem der Lehrplan Passepartout im Schuljahr 2022/23 in den Lehrplan 21 übergeht (vgl. *Erläuterungen im Französisch und Englisch im Kanton Wallis*, Departement für Volkswirtschaft und Bildung des Kantons Wallis, 2018).

Der Lehrplan 21 für Sprachen besteht für Deutsch und die erste und zweite Fremdsprache, wobei jede Sprache separat aufgebaut wird, aber die gemeinsame Struktur es dennoch ermöglicht, Synergien zu nutzen. Während das Erlernen einer zweiten Landessprache wirtschaftlich und gesellschaftlich begründet wird, steigt der Bedarf an kompetenten Englischsprechenden, da Englisch in der internationalen Politik, im internationalen Handel und in den digitalen Medien einen bedeutenden Stellenwert einnimmt. Im Fremdsprachenunterricht werden sprachliche Kompetenzen in den Bereichen Hören, Lesen, Sprechen und Schreiben entwickelt und gefördert. Die Aufteilung der Fertigkeiten in diese Kompetenzbereiche unterstützt einen der Kerngedanken des GER, nämlich die Anerkennung von ungleich ausgeprägten Kompetenzen in einer Sprache (Konsortium ÜGK, 2019, S. 35). Die Schülerinnen und Schüler lernen, sprachliche Fertigkeiten in authentischen Situationen anzuwenden, wobei die Inhalte auf interessanten Themen aus ihrem Alltag basieren. Die Arbeit an Wortschatz, Grammatik und Rechtschreibung dient in erster Linie der kommunikativen Handlung und nicht dem Selbstzweck. Diese Arbeit wird im

Kompetenzbereich «Sprache(n) im Fokus» beschrieben (Departement für Volkswirtschaft und Bildung des Kantons Wallis, 2018, S. 61ff.).

Der Lehrplan 21 legt in den strukturellen und inhaltlichen Hinweisen die erwarteten Grundkompetenzen in Fremdsprachen auf Basis des Gemeinsamen Europäischen Referenzrahmens für Sprachen (GER) fest. Die Mindestanforderungen am Ende der obligatorischen Schulzeit liegen im Schreiben in beiden Fremdsprachen auf Niveau A 2.1 nach dem GER (Departement für Volkswirtschaft und Bildung des Kantons Wallis, 2018, S. 63).

Die Schreibaufgabe, welche dieser Untersuchung zugrunde liegt, ist einerseits im Kompetenzbereich «Schreiben» unter der Rubrik «A: Schriftliche Texte verfassen» und andererseits im Kompetenzbereich «Sprache(n) im Fokus» unter den Rubriken «B: Wortschatz», «D: Grammatik» und «E: Rechtschreibung» angesiedelt. Gemäss Lehrplan 21 werden u.a. folgende Grundkompetenzen in der ersten und zweiten Fremdsprache im Bereich «Schreiben» angestrebt (vgl. Departement für Volkswirtschaft und Bildung des Kantons Wallis, 2018):

FS1F.4.A.1 / FS2E.4.A.1 *Die Schülerinnen und Schüler...*

- c. können mit sprachlichem Support (z.B. Chunks) einfache Geschichten beginnen oder zu Ende führen.

können sich einfache Notizen machen und kurze Mitteilungen verfassen [...]

können mit einfachen Sätzen und Ausdrücken über vertraute Gegenstände und Personen berichten [...]

In den Grundansprüchen im Zyklus 3 wird im Bereich «Sprache(n) im Fokus» u.a. in folgenden Kompetenzstufen gearbeitet:

FS1F.5.B.1 / FS2E.5.B.1 *Die Schülerinnen und Schüler...*

- c. verfügen über ein genügend grosses Repertoire an Wörtern und Wendungen, das ihnen erlaubt, Texte zu allgemeinen Themen zu verfassen [...]

FS1F.5.D.1 / FS2E.5.D.1 *Die Schülerinnen und Schüler...*

- c. können beim [...] Schreiben einfache grammatische Strukturen verwenden, machen dabei aber teilweise noch Fehler [...]

FS1F.5.E.1 / FS2E.5.E.1 *Die Schülerinnen und Schüler...*

- c. können in kürzeren Texten zu vertrauten Themen angemessen korrekt schreiben.

können häufig verwendete Formen korrekt schreiben [...]

2.3 Schreibkompetenzen in der ersten und zweiten Fremdsprache im Vergleich

Verschiedene Forschende gehen davon aus, dass die Verwendung einer angemessenen Wortwahl einen grossen Einfluss auf die Textqualität hat (z. B. S. A. Crossley & McNamara, 2011; Olinghouse & Wilson, 2013) und suchen daher nach einer Möglichkeit, den in Texten verwendeten Wortschatz objektiv zu bewerten. Automatisierte Messmethoden, die es ermöglichen, den Wortschatz in Arbeiten von Lernenden zu bewerten, sind von grosser Bedeutung für diejenigen, die an der Unterscheidung zwischen Texten von Lernenden mit unterschiedlichen Sprachniveaus interessiert sind, wie sie im GER gemessen werden (Treffers-Daller et al., 2016, S. 2). Es gibt jedoch Herausforderungen bei der Verwendung des GER-Rahmens. Ein Schüler oder eine Schülerin kann sich auf unterschiedlichen Niveaus in Bezug auf Wortschatz, grammatikalische Genauigkeit und phonologische Kontrolle befinden. Ausserdem gibt es wenig empirische Evidenz für die angeblichen Zusammenhänge zwischen den GER-Stufen und standardisierten Prüfungen (z. B. Alderson, 2007; Hulstijn, 2007).

Treffers-Daller et al. (2016) berichten in ihrer Studie, wie Messungen der lexikalischen Vielfalt bei der Unterscheidung zwischen den Niveaus des GER nützlich sein können. Dabei wird hervorgehoben, dass Masse zur Messung der lexikalischen Diversität für die automatisierte Überprüfung des Wortschatzes von Lernenden in Schreibaufgaben von grossem Nutzen sein können, dass diese jedoch nicht in der Lage sind, zwischen den Stufen des GER zu unterscheiden. Es ist jedoch unrealistisch, dass ein einziges Mass alle diese Niveaus nach dem GER unterscheiden kann, da sie sich teilweise überschneiden.

Peyer et al. (2016b) haben im Auftrag der Bildungsdirektoren-Konferenz Zentralschweiz die Fremdsprachenkompetenz in der Bildungsregion Zentralschweiz evaluiert. Die Studie

fokussierte sich auf die beiden Schulfremdsprachen Französisch und Englisch in der 8. Klasse (10H) und untersuchte, ob die Lehrplanziele erreicht wurden.

Für die vorliegende Arbeit sind die Ergebnisse im Bereich Schreiben relevant. Laut Studienergebnissen erreicht eine Minderheit der Lernenden im Bereich Schreiben im Französischunterricht das erforderliche Niveau A2.1 am Ende der 8. Klasse (10H) (Peyer et al., 2016b, S. 43). Im Gegensatz dazu erreichen rund zwei Drittel der Lernenden im Englischunterricht die Lehrplanziele (Peyer et al., 2016b, S. 47). Die Studie zeigt deutliche Unterschiede zwischen den Sprachkompetenzen in der ersten und zweiten Fremdsprache (Peyer et al., 2016b, S. 53). Es ist zu beachten, dass in der BKZ-Region Englisch als erste und Französisch als zweite Fremdsprache unterrichtet werden, was sich auch in den durchschnittlichen Jahreswochenlektionen zeigt. Die Lernenden hatten im Durchschnitt 15,5 Jahreswochenlektionen Englischunterricht und nur 11,1 Jahreswochenlektionen Französischunterricht, wobei die Anzahl der Jahreswochenlektionen teilweise stark variiert. Peyer et al. (2016b, S. 50) gehen davon aus, dass es unwahrscheinlich ist, dass sich die Kompetenzen im Fach Französisch, welches in der untersuchten Region die zweite Fremdsprache ist, bis zum Ende der obligatorischen Schulzeit den Kompetenzen im Fach Englisch, welches die erste Fremdsprache ist, angleichen werden.

2.4 Lexikalischer Reichtum – Lexikalische Vielfalt

Obwohl der lexikalische Reichtum (engl. *lexical richness*) mindestens seit der ersten Hälfte des 20. Jahrhunderts untersucht wird, bleibt er ein komplexes Konzept (Bonvin & Lambelet, 2019, S. 65). Einerseits gibt es Unterschiede in der Terminologie und in der Bedeutung des Begriffs, andererseits gibt es kein allgemein gültiges Mass für lexikalischen Reichtum, welches die verschiedenen Komponenten, die der Begriff «lexikalischer Reichtum» abdeckt, berücksichtigt und den Vergleich aller Arten von Texten ermöglicht.

In den folgenden Abschnitten werden für die Untersuchung relevante Begriffe definiert und auf die verschiedenen Komponenten, die in bisherigen Studien zur Messung des lexikalischen Reichtums verwendet wurden, eingegangen. Es sei an dieser Stelle erwähnt, dass nachfolgend einige relevante Studien und Tendenzen in Zusammenhang mit dem lexikalischen Reichtum zusammengefasst werden, die Liste aber bei weitem nicht abschliessend ist.

2.4.1 Terminologie

Jarvis (2013a, S. 4) zeigt auf, dass die heutigen Vorstellungen von lexikalischer Vielfalt (engl. *lexical diversity*, franz. *diversité lexicale*) und lexikalem Reichtum (engl. *lexical richness*, franz. *richesse lexicale*) aus einer ziemlich intensiven Periode lexikalischer Grundlagenforschung in den Bereichen Linguistik, Psychologie und literaturstatistische Analyse stammen, die sich von etwa 1935 bis 1944 erstreckt hat und sich auch auf andere verwandte Bereiche ausgeweitet haben, wie beispielsweise den Erst- und Zweitspracherwerb oder die Zwei- und Mehrsprachigkeit.

Der Begriff des lexikalischen Reichtums wird oft als Überbegriff für eine Reihe von Messungen verwendet (z.B. Bonvin et al., 2018, S. 136; Bonvin & Lambelet, 2019, S. 66), die je nach Autor variieren (z.B. Jarvis, 2017; Read, 2000) und gehen von der lexikalischen Vielfalt als bekanntestes und ältestes Mass des lexikalischen Reichtums aus. Die lexikalische Vielfalt wird bei Bonvin et al. (2018, S. 137) auch als Wortschatzvarianz bezeichnet. Der Begriff des lexikalischen Reichtums hat sich im Laufe der Zeit verändert (Jarvis, 2013a, S. 48).

Für Yule (1944, S. 83) bezieht sich der Begriff des lexikalischen Reichtums (oder: *vocabulary richness*) auf die Anzahl der Wörter im mentalen Lexikon einer Person. Im Gegensatz dazu scheint beispielsweise Honoré (1979) den Begriff des lexikalischen Reichtums mit dem Ausdruck der lexikalischen Vielfalt gleichgesetzt zu haben, während Engber (1995), Read (2000), Bonvin & Lambelet (2019) und andere ihn als übergeordnete Kategorie begreifen, um die lexikalische Vielfalt als eines der vielen lexikalischen Merkmale guten Schreibens zu umfassen. So verwenden beispielsweise Engber (1995) oder Read (2000) den Terminus der lexikalischen Vielfalt, um sich lediglich auf den Diversitätsaspekt zu beziehen. Mit anderen Worten, die Bedeutung des lexikalischen Reichtums hat sich von einer sehr spezifischen Bedeutung, die sich nicht mit der der lexikalischen Vielfalt überschneidet, zu einer Bedeutung erweitert, die über die lexikalische Vielfalt hinausgeht. Einige Forschende, wie beispielsweise Vanhove et al. (2019) oder Jarvis (2013a), verwenden den Begriff der «lexikalischen Vielfalt» anstelle des Begriffs «lexikalischer Reichtum», um sich auf die Menge, den Umfang und die Vielfalt des in einem Text verwendeten Vokabulars zu beziehen.

Eine weitere Komponente, die u.a. gemäss Bonvin & Lambelet (2019, S. 66) zum lexikalischen Reichtum beiträgt, ist die lexikalische Elaboriertheit (engl. *lexical sophistication*, franz.

sophistication lexicale), welche durch die Verwendung von fortgeschrittenen, schwierigen oder seltenen Wörtern definiert wird. Da die Klassifizierung der Wörter von den Forschenden und den Untersuchungsgruppen abhängig ist, wie beispielsweise bei Laufer (1994, S. 22), gibt es keinen allgemein gültigen Konsens über die Definition der *lexical sophistication*. Wie in Bonvin & Lambelet (2019, S. 66) beschrieben, wird die Komplexität eines Textes oft durch die Seltenheit von Wörtern, gemessen an ihrer Häufigkeit in der repräsentativen Gruppe der betreffenden Sprache, bestimmt.

Obwohl Änderungen der Begrifflichkeiten oft mit theoretischen Entwicklungen in einem Fachgebiet zusammenfallen, können diese auftreten, wenn Begriffe nicht in Definitionen verankert sind. Die Entwicklung der lexikalischen Statistik hat erhebliche Fortschritte bei der Geschwindigkeit und Vergleichbarkeit der Bewertung des lexikalischen Reichtums gesprochener und geschriebener Texte ermöglicht. Die verschiedenen Aspekte des lexikalischen Reichtums können jedoch durch verschiedene Algorithmen gemessen werden, wovon einige im Kapitel 2.4.2 diskutiert werden. Das schwerwiegendste und hartnäckigste Problem in der lexikalischen Diversitätsforschung scheint wohl zu sein, dass Messungen zwar immer ausgefeilter werden, sich in Bezug auf das untersuchte Konzept jedoch als zunehmend komplexer und weniger transparent erweisen (Bonvin & Lambelet, 2019, S. 67; Jarvis, 2013a, S. 48). In dieser Untersuchung wird der Begriff der lexikalischen Vielfalt als «the variety of word use that can be found in a person's speech or writing» verwendet (Jarvis, 2013a, S. 25).

2.4.2 Operationalisierung der lexikalischen Vielfalt

Gemäss Bonvin & Lambelet (2019, S. 65) stellen sich bei der Operationalisierung des Konzepts des lexikalischen Reichtums in der Literatur zwei Hauptherausforderungen dar. Erstens werden die zur automatischen Messung entwickelten Algorithmen von der Textlänge beeinflusst, was es schwierig macht, Texte unterschiedlicher Länge zu vergleichen. Zweitens unterscheiden sich lexikalisch-syntaktische Merkmale in verschiedenen Sprachen, so dass es unmöglich ist, den lexikalischen Reichtum von Texten in verschiedenen Sprachen direkt zu vergleichen. In den folgenden Abschnitten wird aufgezeigt, mit welchen Ansätzen verschiedene Autoren das Konzept des lexikalischen Reichtums untersucht haben, wie sich diese Konzepte entwickelt haben und an welchen Methoden sich gegenwärtige Forschungsarbeiten orientieren.

1935 beschreibt der amerikanische Linguist George Zipf das sogenannte Zipfsche Gesetz. Dieses Potenzgesetz wird seither auf viele Phänomene in der physikalischen Welt ausserhalb der Linguistik angewendet. Es besagt, dass das häufigste Wort in einem Text etwa zehn Prozent des Gesamttextes ausmacht, das zweithäufigste etwa fünf Prozent und das dritthäufigste etwa einen Dreissigstel (Zipf, 1935). Die direkteste Folge dieses Gesetzes ist, dass Wortrang und Worthäufigkeit in umgekehrter Beziehung zueinanderstehen. Dieses Prinzip hat jedoch auch andere Implikationen. 1937 beschreibt Zipf die Wirkung einzelner Wörter auf die Anzahl ihrer Wiederholungen im Text (Zipf, 1937). Er argumentiert, dass die Wortwiederholungsrate vorhersagbar ist. Beispielsweise wird das am häufigsten verwendete Wort im Text in relativ regelmässigen Abständen wiederholt, etwa einmal alle 10 Wörter, und das am zweithäufigsten verwendete Wort wird etwa alle 20 Wörter wiederholt (Zipf, 1937).

Auch der amerikanische Psychologe John B. Carroll (1938) prägt, ausgehend von der Vorstellung der durchschnittlichen Wortwiederholungsrate von Zipf (Carroll, 1938, S. 380), den Begriff der lexikalischen Vielfalt. Carrolls eigene Definition der lexikalischen Vielfalt besagt, dass es sich um den relativen Grad der Wiederholung oder die relative Vielfalt des Vokabulars handelt (Carroll, 1938, S. 379). Es werden zwei Begriffe verwendet, damit dieses Problem besser verstanden werden kann: Type und Token. Die beiden Begriffe beziehen sich, wenn sie auf den Wortschatz angewandt werden, auf unterschiedliche Arten der Quantifizierung von Worthäufigkeiten in einer Sprachprobe (Jarvis, 2013b, S. 15). Wird ein Text länger, steigt die Anzahl der Token linear, aber nicht die Anzahl der Types, weil die Wahrscheinlichkeit eines neuen Type mit zunehmender Textlänge abnimmt, da Texte ohne Wiederholungen von gewissen Type nicht sinnvoll sein können, was dazu führt, dass der Wert der Vielfalt eines Textes mit zunehmender Länge abnimmt (McCarthy, 2005, S. 382).

Einerseits verdeutlicht Carrolls Definition, dass die lexikalische Vielfalt durch die Anzahl der verschiedenen Vorkommnistypen (engl. *type*) in einem Text und der Wiederholung in der Anzahl (engl. *token*) der Vorkommnisse beeinflusst wird. Andererseits ist Carroll mit seiner Definition der Überzeugung, dass die Anzahl der Types und ihre Wiederholungshäufigkeit mit der Textlänge relativiert werden kann, sodass Texte unterschiedlicher Länge auf derselben Diversitätsskala verglichen werden können. Seine empirischen Untersuchungen zeigen aber, dass weder die Wiederholungsrate noch der von Zipf vorgeschlagene Zusammenhang zwischen Rang und Häufigkeit für unterschiedliche Stichprobenumfänge konstant bleiben. Um

dieses Problem zu lösen, führt Carroll eine unzureichende Gleichung für die Vielfalt ein, die zu erfassen versucht, wie die Wiederholungsrate bei unterschiedlichen Stichprobengrößen variiert (Jarvis, 2013a, S. 45).

2.4.2.1 Type-Token-Verhältnis (TTR)

Mit dem Wunsch das Problem der Stichprobengröße zu lösen, schlägt Johnson (1944) die Type-Token-Ratio (TTR) als potenziell nützlichen Indikator für das vor, was er Wortschatzflexibilität oder -variabilität bezeichnet. Dieses Mass wird bereits von Johnson (1944) erwähnt, wird aber auch häufig Templin (1957) zugeschrieben. Die TTR besteht also aus dem Verhältnis zwischen der Anzahl der Vorkommnistypen (*types*) dividiert durch die Anzahl der Vorkommnisse (*token*). Während die TTR intuitiv nützlich ist, reagiert sie empfindlich auf die Textgröße und ist beim Vergleichen von Texten unterschiedlicher Länge von geringem Nutzen, denn je länger ein Text ist, desto unwahrscheinlicher erscheinen neue Wörter und desto wahrscheinlicher werden Wörter wiederholt (McCarthy & Jarvis, 2007, S. 60). Folglich sinkt zwangsläufig der TTR-Wert mit zunehmender Textlänge (McCarthy & Jarvis, 2013, S. 48f.). Dies liegt daran, dass die Anzahl der Tokens schneller wächst als die Anzahl der verschiedenen Types. Da die Standard-TTR dazu neigt, mit zunehmender Länge eines Textes zu sinken, werden längere Aufsätze benachteiligt (Mellor, 2011, S. 3). Um dieses Problem anzugehen, wurden in den letzten Jahrzehnten viele andere Metriken entwickelt, die weniger intuitiv, aber widerstandsfähiger gegenüber Änderungen der Größe des analysierten Textes sind (z.B. Carroll, 1967; Guiraud, 1954; Herdan, 1964; Honoré, 1979; Jarvis, 2002; Malvern et al., 2004; McCarthy & Jarvis, 2007; Sichel, 1975, 1986; Tweedie & Baayen, 1998).

2.4.2.2 Konstante von Yule

Ein Lösungsansatz findet sich in einer von manchen Forschenden (Malvern, D.D & Richards, B.J., 1997) als uninterpretierbar bezeichneten Formel von Yule (1944) wieder. Yule (1944) berechnet anhand der Formel, mit welcher Wahrscheinlichkeit bei der zufälligen Auswahl von Wörtern aus einem Text dasselbe Wort zweimal hintereinander gewählt wird (Baayen, 2001, S. 25). Da eine höhere Wahrscheinlichkeit, dass dasselbe Wort zweimal hintereinander gewählt wird, im Wesentlichen einer höheren Wiederholungsrate entspricht, kann die charakteristische Konstante von Yule (d. h. der Koeffizient in seiner Formel) als Index für die Rate der Wortwiederholungen in einem Text verwendet werden. Obwohl nach Baayen (2001, S. 211) u.a. die Messwerte von Yule (1944) wirklich theoretisch unabhängig vom

Stichprobenumfang zu sein scheinen, erweisen sie sich in der Praxis dennoch als nicht unabhängig von der Stichprobengrösse (Jarvis, 2013a, S. 47).

2.4.2.3 Guiraud-Index

Der Guiraud Index (Guiraud, 1954) wird berechnet, indem die Anzahl der unterschiedlichen Wörter (*types*) durch die Quadratwurzel der Anzahl der Wörter in einem Text (*tokens*) dividiert wird ($\frac{\text{Anzahl unterschiedlicher Wörter}}{\sqrt{\text{Anzahl Wörter}}}$). Gemäss Malvern et al. (2004) steigt der Guiraud-Index mit zunehmender Aufsatzlänge in den ersten paar hundert Wörtern an, nimmt dann aber mit zunehmender Aufsatzlänge stetig ab. Obwohl es einen Zusammenhang zwischen der Anzahl der Types und der Textlänge gibt, bleibt der Guiraud-Index bei Texten mit mehreren hundert Wörtern stabil, wie Daller (2010, S. 12) betont. Greber (2018, S. 122) zeigt in der Gegenüberstellung der herkömmlichen TTR und des Guiraud-Index auf, dass zwar der Wert für das gleiche Type-Token-Verhältnis mit zunehmender Textlänge nicht variiert, im Vergleich dazu der Guiraud-Index bei zunehmendem Textumfang aber kontinuierlich ansteigt. Dies kann am Beispiel illustriert werden, dass bei 20 Types und 30 Tokens die TTR bei 0.67 liegt und der Guiraud-Index bei 3.65, während bei 80 Types und 120 Tokens der TTR-Wert nach wie vor bei 0.67 liegt und der Guiraud-Index auf 7.3 ansteigt (Greber, 2018, S. 122).

2.4.2.4 Mean Segmental Type-Token Ratio (MSTTR)

Johnson's (1944) Lösung für das Problem der Stichprobengrössenabhängigkeit besteht darin, die TTR unter Verwendung einer Standardanzahl von Token für jeden Text (z. B. die ersten 200 Wörter) zu berechnen oder alternativ den Durchschnitt mehrerer gleich grosser Teilstichproben eines Textes zu ermitteln. Johnson nennt diese letztere Alternative Mean Segmental Type-Token Ratio (MSTTR). Seine Lösung scheint die Probleme der Stichprobengrössenabhängigkeit teilweise zu überwinden (vgl. Covington & McFall, 2010; Malvern et al., 2004).

2.4.2.5 Measure of textual lexical diversity (MTLD)

Mit der *measure of textual lexical diversity* (MTLD) liefert McCarthy (2005) ein Mass, welches wie Johnson's (1944) MSTTR über weite Bereiche der Stichprobengrösse relativ konstant zu sein scheint (vgl. Covington & McFall, 2010; Malvern et al., 2004; McCarthy & Jarvis, 2010). Mit der MTLD wird die TTR nach jedem Wort in der Stichprobe gemessen, bis sie den Wert 0.72 erreicht. Die Messung der TTR wird mit dem nächsten Token fortgesetzt, bis das letzte

Token der Probe berücksichtigt wird. Die Länge des Textes wird dann durch die Gesamtzahl der TTR von 0.72 geteilt (Fergadiotis et al., 2015, S. 3; McCarthy & Jarvis, 2010, S. 385). In entgegengesetzter Richtung, d.h. vom letzten Wort bis zum ersten Wort, wird eine weitere MTLD-Messung vorgenommen. Im Anschluss wird der MTLD-Index gemäss Koizumi (2012, S. 62ff.) aus dem arithmetischen Mittel der vorwärts und rückwärts gemessenen MTLD-Werte bestimmt.

Das MTLD-Mass hat mehrere Vorteile, wie aus verschiedenen Studien ersichtlich ist. Einige Untersuchungen haben gezeigt, dass es unempfindlicher gegenüber Variationen der Textlänge ist als andere Masse, wie z.B. D oder HD-D (Fergadiotis et al., 2013; Treffers-Daller, 2013), die weiter unten aufgeführt werden. Ein weiterer Vorteil, der sich in einigen Studien manifestiert, ist, dass das MTLD-Mass keine Verzerrungen bei Textproben mit 100 bis 2000 Tokens aufweist (S. Crossley et al., 2009; McCarthy, 2005; zitiert in Treffers-Daller, 2013, S. 82).

Wie Koizumi (2012, S. 67) jedoch feststellt, ist die MTLD trotzdem empfindlich für Stichprobeneffekte bei Texten mit 50 bis 100 Tokens, 100 bis 200 Tokens und 50 bis 200 Tokens. Aufgrund dieser Erkenntnisse empfiehlt Koizumi (2012, S. 67), dass MTLD bei Texten mit mindestens 100 Tokens und, wenn möglich, mit einer maximalen Differenz von 50 Tokens zwischen den Texten, eingesetzt werden sollte.

2.4.2.6 D

Mit dem Mass D, welches aus der Entwicklung der Computerlinguistik hervorgeht und eine Anpassung des TTR vorschlägt, soll die Geschwindigkeit berechnet werden, mit welcher die TTR in einer Sprachprobe abnimmt. Der D-Wert ist in der L1- und L2-Forschung weit verbreitet (Malvern et al., 2004, S. 55f.). Dabei stützt sich dieses Mass auf eine Stichprobenreihe, in welcher die TTR für jeweils 100 Stichproben zwischen 35 und 50 Token berechnet wird. Das Ergebnis ist eine Annäherung an den Wert mit allen möglichen Stichproben (McCarthy & Jarvis, 2010, S. 383). Duran et al. (2004, S. 220ff.), welche D mit kurzen Texten in der Erst- und Fremdsprache getestet haben, kommen zu dem Schluss, dass es ein guter Prädiktor für die Sprachentwicklung ist. Allerdings reagiert auch dieses Mass empfindlich auf die Textlänge (Fergadiotis et al., 2015), was insbesondere bei kurzen Stichproben mit weniger als 150 Token festgestellt werden kann (Koizumi, 2012, S. 67).

2.4.2.7 HD-D

HD-D wird von McCarthy & Jarvis (2007, 2010) als Alternative zu D vorgeschlagen. Insbesondere bestimmt HD-D die Wahrscheinlichkeit jedes Type im Text anhand einer zufälligen Teilstichprobe von 42 Tokens, d.h. es wird für jeden lexikalischen Type in einem Text die Wahrscheinlichkeit berechnet, eines seiner Token in einer Zufallsstichprobe von 42 Wörtern aus dem Text zu finden. Der Index der Lexikalischen Vielfalt wird dann berechnet, indem die Wahrscheinlichkeiten jedes Types, der im Text vorkommt, summiert wird (McCarthy & Jarvis, 2010, S. 383).

Obwohl schlussendlich das Problem beim Nichtvorhandensein der Konstruktvalidität eines gut entwickelten theoretischen Modells der lexikalischen Vielfalt liegt (Jarvis, 2013a, S. 48), haben verschiedene Studien gezeigt, dass Indikatoren der Wortschatzqualität selbst bei unvollkommener Messung nützliche Prädiktoren sind für andere wichtige Konstrukte wie beispielsweise die Sprachkenntnisse, die Komplexität der Sprache oder des Wortschatzes (z. B. Berman et al., 2011; Crossley et al., 2011; Housen et al., 2011; Sauro & Smith, 2010; Yu, 2010). Damit zeigt Jarvis (2013a, S. 47) auf, dass Indizes der lexikalischen Vielfalt nützlich sind, auch wenn Sprachforscher die Frage vernachlässigt haben, was sie eigentlich messen. Der Einfluss der Textlänge bleibt jedoch ein bedeutender Faktor in der Spracherwerbsforschung, die sich mit viel kürzeren Texten befasst (Duran et al., 2004, S. 222). Ein Mass für lexikalische Vielfalt, das ermöglicht, Texte unterschiedlicher Länge miteinander zu vergleichen, ist eine Art «Heiliger Gral» in der Forschung über lexikalische Vielfalt (Treffers-Daller et al., 2016, S. 4). Die fokussierte Konzentration auf die Lösung des Textlängenproblems hat die Sprachforschenden daran gehindert, von wichtigen Fortschritten in anderen Bereichen zu profitieren (Jarvis, 2013a, S. 48).

2.4.2.8 Lexikalischer Reichtum als Wahrnehmungsphänomen

Ausserdem schlägt Jarvis (2013a, S. 48) vor, lexikalischen Reichtum als ein Wahrnehmungsphänomen zu behandeln (vgl. Vanhove et al., 2019, S. 500), weshalb zukünftige Indizes der lexikalischen Vielfalt nicht auf ihre Fähigkeit getestet werden sollten, Sampling-Effekte und andere Variablen, wie beispielsweise Leistungsbewertungen vorherzusagen oder zu überwinden, sondern eher durch ihre Fähigkeit, die Urteile lexikalischer Vielfalt menschlicher Bewerter vorherzusagen. Zunächst wird beobachtet, dass menschliche Bewerterinnen und Bewerter bei der Beurteilung der lexikalischen Fähigkeiten

der Lernenden ein hohes Mass an Konsistenz erreichen können, auch mit nur geringem Training (Jarvis, 2013b, S. 22). Die Untersuchungen haben gezeigt, dass lexikalische Kompetenzurteile von menschlichen Gutachterinnen und Gutachtern zu 60% genau vorhergesagt werden können, indem eine angemessen gewichtete und sorgfältig ausgewählte Kombination aus objektiven lexikalischen Massen verwendet wird. Diese Ergebnisse geben Anlass zur Zuversicht, dass gut entwickelte objektive Masse auch menschliche Urteile über die lexikalische Vielfalt vorhersagen können (Jarvis, 2013b, S. 22).

Polio (2012, S. 145f.) geht davon aus, dass die Sprachkompetenz in der Forschung als eine Kombination aus Komplexität, Genauigkeit und Geläufigkeit untersucht wird und Indizes der lexikalischen Vielfalt oft in Verbindung mit anderen Komplexitätsmassen verwendet werden. Gemäss Bonvin & Lambelet (2019) gibt es nur wenige Studien, die sich auf die menschliche Bewertung von lexikalischem Reichtum beziehen (z.B. Bonvin & Lambelet, 2017, 2019; Jarvis, 2017, 2013a, 2013b; Vanhove et al., 2019).

Jarvis (2013a, S. 48ff.) weist ebenfalls darauf hin, dass der Text als Ganzes bewertet werden sollte und eine qualitativere Betrachtung seiner lexikalischen Vielfalt in Erwägung gezogen werden sollte. Damit der lexikalische Reichtum als Wahrnehmungsphänomen betrachtet werden kann, plädiert Jarvis (2013a, S. 52) für den Einsatz menschlicher Bewerterinnen und Bewerter, welcher auch mit einigen Komplikationen einhergeht, beispielsweise bezüglich der Schulung von Bewertenden. Jarvis (2013a, S. 52) geht davon aus, dass die menschliche Bewertung der lexikalischen Vielfalt von verschiedenen miteinander verknüpften Faktoren beeinflusst wird und nimmt die von Ökologen beschriebenen Aspekte der Vielfalt als Ausgangspunkt für die sieben Merkmale der lexikalischen Vielfalt, welche es zu berücksichtigen gilt. Um ein geeignetes Mass für dieses subjektive Konstrukt in Bezug auf die Wahrnehmung der lexikalischen Vielfalt zu erhalten, müssen die Dimensionen der lexikalischen Vielfalt bestimmt und Masse dafür entwickelt werden, sodass diese kombiniert und mit den tatsächlichen menschlichen Wahrnehmungen abgeglichen werden können (Jarvis, 2013b, S. 22). Daraus werden folgende Dimensionen abgeleitet:

1. **Volumen** (Size: vgl. Jarvis, 2013a, S. 52; Volume: vgl. Jarvis, 2013b, S. 22; Vanhove et al., 2019, S. 501)

Das Volumen bezieht sich auf die Textlänge, d.h. auf die Anzahl der Token. Die Textlänge wird in Studien zur lexikalischen Vielfalt oft als Störfaktor angesehen. Dies

kann gemäss Koizumi (2012) oder Malvern (2008) darauf zurückgeführt werden, dass Masse der Worthäufigkeit nicht sinnvoll über Texte unterschiedlicher Länge miteinander verglichen werden können. Wie Jarvis (2017, 2013b, 2013a) beschäftigen sich auch Vanhove et al. (2019) damit, wie Leser die lexikalische Vielfalt wahrnehmen. Die Einbeziehung des Umfangs eines Textes spiegelt die Möglichkeit wider, dass Leserinnen und Leser den lexikalischen Reichtum der verglichenen Texte je nach Länge der Texte unterschiedlich wahrnehmen (Vanhove et al., 2019, S. 501). Zu den berechneten Werten bezüglich Volumen gehören u.a. die Anzahl der Token (vgl. auch Jarvis, 2013b, S. 29), der Types, der eindeutigen Lemmata oder die Anzahl der Sätze (Vanhove et al., 2019, S. 508). Jarvis (2013b, S. 29) deutet zudem darauf hin, dass es eine Herausforderung ist, zu definieren, was ein Wort ist (z.B. zusammengesetzte Wörter) und dass dieser Massstab ebenfalls sprachübergreifend funktionieren muss.

2. **Reichtum** (*Richness*: vgl. Jarvis, 2013a, S. 52; *Variability*: vgl. Jarvis, 2013b, S. 22; Vanhove et al., 2019, S. 501)

Der Reichtum wird durch die Anzahl der Types bestimmt (Jarvis, 2013a, S. 52). Das bekannteste Messinstrument ist die TTR (vgl. Kapitel 2.4.2.1), d.h. das Verhältnis der Anzahl eindeutiger Wörter im Text, zur Gesamtzahl der Wörter. Da die meisten Masse (vgl. Kapitel 2.4) systematisch mit der Textlänge verbunden sind, überschneiden sich die Dimensionen des Volumens und des Reichtums (Vanhove et al., 2019, S. 501). Vanhove et al. (2019, S. 508) zählen zu den Indikatoren für die Variabilität ausschliesslich Masse, die auf der Anzahl der in den Texten vorkommenden Types und Token basieren, wie beispielsweise die TTR, den Guiraud Index, Yules K, HD-D oder die MTLD. Jarvis (2013b, S. 28) plädiert in Bezug auf den Reichtum auf das Mass der MTLD, da diese das einzige Mass ist, welches nicht in Abhängigkeit zum Volumen oder der Gleichmässigkeit steht.

3. **Effektive Anzahl der Types** (*effective number of types*: vgl. Jarvis, 2013a, S. 52)
4. **Gleichmässigkeit** (*evenness*: vgl. Jarvis, 2013a, S. 52, 2013b, S. 23; Vanhove et al., 2019, S. 501)

Die Gleichmässigkeit definiert den Grad, in welchem die Token gleichmässig über die Typen verteilt sind (Jarvis, 2013a, S. 52). Damit sollen Unterschiede im Ausmass erfasst werden, in dem Token verschiedener Types zum Text beitragen, d.h. wie regelmässig

ein Type im Vergleich zu anderen Types auftritt (Vanhove et al., 2019, S. 501). Als erste Operationalisierung der Gleichmässigkeit verwendet Jarvis (2013b, S. 29) die Standardabweichung der Anzahl der Token pro Type.

5. **Disparität** (*disparity*: vgl. Jarvis, 2013a, S. 52; Vanhove et al., 2019, S. 502)

Mit der Disparität wird der Anteil der Wörter in einem Text bezeichnet, welche formal oder semantisch verwandt sind (Jarvis, 2013a, S. 52, 2013b, S. 25). Obschon die formale Disparität bei Jarvis (2013a) nicht operationalisiert wird, berechnen Vanhove et al. (2019, S. 502) die Levenshtein-Distanzen, wobei eine höhere mittlere Distanz auf eine grössere formale Disparität des Textes hinweist.

6. **Seltenheit** (*importance*: vgl. Jarvis, 2013a, S. 52; *rarity*: vgl. Vanhove et al., 2019, S. 501f.)

Mit der Seltenheit, bzw. Wichtigkeit oder Elaboriertheit (engl. *lexical sophistication*), ist die Verwendung weniger gebräuchlicher, seltener Wörter gemeint. Es ist eines von vielen lose miteinander verbundenen Massen, mit welchem die relative Häufigkeit berechnet wird, mit der die Wörter in einem Text in der Sprache als Ganzes vorkommen, womit beispielsweise ein Bezug zu einem repräsentativen Korpus hergestellt wird (Jarvis, 2013a, S. 52, 2013b, S. 29). Als womöglich einfachstes Mass für die Berechnung der Seltenheit bezeichnet Jarvis (2013b, S. 29) die Berechnung des mittleren Rangs für alle Wörter eines Textes, indem die lemmatisierte Rangfolge jedes Wortes im Referenzkorpus identifiziert wird. In Anlehnung an Laufer & Nation (1995) nennen Jarvis (2013b, S. 30) und Vanhove et al. (2019, S. 509) auch die Berechnung von *frequency bands* als weitere Option. Mit Bezugnahme zu Kyle & Crossley (2015) schlagen Vanhove et al. (2019, S. 509) ebenfalls die Berechnung der mittleren Korpus Häufigkeit oder die Berechnung der TTR und des Guiraud-Index anhand der relativ seltenen Types gemäss Daller et al. (2003) vor.

7. **Dispersion** (*dispersion*: vgl. Jarvis, 2013a, S. 52; Vanhove et al., 2019, S. 502)

Damit ist der durchschnittliche Abstand zwischen Token desselben Types gemeint (Jarvis, 2013a, S. 52). Jarvis (2013b, S. 25) geht davon aus, dass eine grössere Streuung der Types zu einem höheren Niveau der wahrgenommenen lexikalischen Vielfalt führt. Vanhove et al. (2019, S. 509) haben in Anlehnung an Jarvis (2013b) einen Dispersionsindex berechnet, wobei höhere Werte auf eine geringere Streuung hinweisen.

Bonvin et al. (2018, S. 136) haben in ihrer Arbeit die lexikalische Vielfalt und Korpusfrequenz als Hauptkomponenten des sogenannten lexikalischen Reichtums untersucht. Diese Grössen werden in den meisten Studien zum Erlernen eines Wortschatzes in einer zweiten Sprache angewendet. Neben der lexikalischen Vielfalt können auch weitere Faktoren wie beispielsweise der Anteil an lexikalischen Fehlern oder die Nutzung von Kollokationen bei der Messung des lexikalischen Reichtums berücksichtigt werden (z.B. Read, 2000, S. 200f.).

Da in dem vorliegenden Korpus nur wenige Kollokationen vorhanden sind und die Anzahl von Fehlern nicht eindeutig lexikalischer oder grammatikalischer Natur sind, konzentriert sich die vorliegende Arbeit ausschliesslich auf die lexikalische Vielfalt. Obwohl die Probleme bezüglich der Textlänge noch nicht gelöst sind, wird weitere Forschung benötigt, um wichtige Informationen liefern zu können. Daher werden in dieser Analyse nur ausgewählte Indizes für die lexikalische Vielfalt herangezogen.

Zusätzlich zu den metrischen Berechnungen werden in der Analyse auch Bewertungen von Gutachterinnen und Gutachtern zur lexikalischen Vielfalt der Texte der Lernenden herangezogen. Aufgrund der typologischen Unterschiede der Sprachen können diese trotz der Lemmatisierung nicht direkt miteinander verglichen werden (Treffers-Daller et al., 2016, S. 11). Daher werden für beide Fremdsprachen Referenztexte herangezogen, mit welchen die Abweichungen der Aufsätze der Schülerinnen und Schüler berechnet werden können. Diese Unterschiede werden anschliessend erfasst und ausgewertet.

Nachdem in den vorherigen Abschnitten die theoretischen Grundlagen zur Messung der lexikalischen Vielfalt erläutert wurden, werden im empirischen Teil der Arbeit die gewonnenen Erkenntnisse auf die Analyse von konkreten Texten von Lernenden der ersten Fremdsprache Französisch und der zweiten Fremdsprache Englisch angewendet. Hierbei sollen die Indizes TTR, MTLD und Guiraud-Index herangezogen werden, um die lexikalische Vielfalt der Texte der Lernenden zu bestimmen. Zudem werden die Bewertungen der Gutachterinnen und Gutachter zur lexikalischen Vielfalt der Texte einbezogen, um ein umfassenderes Bild der lexikalischen Fähigkeiten der Lernenden zu erhalten. Im Folgenden sollen die Ergebnisse der Analyse präsentiert und diskutiert werden.

3 Empirischer Teil

In diesem Kapitel wird die Konzeption der vorliegenden Arbeit beschrieben. In einem ersten Teil wird die Forschungsfrage formuliert, aus der anschliessend die Hypothesen abgeleitet werden. Danach folgt eine Einbettung in den schulischen Kontext, eine detaillierte Beschreibung der Testinstrumente und der Probanden, sowie der ausgewählten Verfahren. Abschliessend werden die Datenerhebung und die Datenanalyse beschrieben.

3.1 Forschungsfrage

Ausgehend von den theoretischen Grundlagen und dem bisherigen Forschungsstand im Bereich der lexikalischen Diversitätsforschung wird mit dieser Forschungsarbeit untersucht, wie sich die lexikalische Vielfalt in schriftlichen Texten von Schülerinnen und Schülern am Ende der obligatorischen Schulzeit in der ersten Fremdsprache Französisch im Vergleich zur zweiten Fremdsprache Englisch im deutschsprachigen Teil des Kantons Wallis unterscheidet. Die Arbeit zielt darauf ab, zu untersuchen, ob die Erwartungen gemäss Lehrplan 21 in Bezug auf die Vergleichbarkeit der schriftlichen Kompetenzen im Hinblick auf die lexikalische Vielfalt in den beiden Fremdsprachen Französisch (L2) und Englisch (L3) am Ende der obligatorischen Schulzeit erfüllt werden. Die Forschungsfrage lautet folgendermassen:

Wie unterscheidet sich die lexikalische Vielfalt in Texten von Schülerinnen und Schülern am Ende der obligatorischen Schulzeit in der ersten Fremdsprache Französisch im Vergleich zur zweiten Fremdsprache Englisch im deutschsprachigen Teil des Kantons Wallis?

3.2 Hypothesen

Die Studienergebnisse von Peyer et al. (2016b) zeigen, dass nur eine Minderheit der Lernenden im Bereich Schreiben im Französischunterricht das erforderliche Niveau A2.1 in der 8. Klasse (10H) erreicht, welches gemäss Lehrplan 21 (Departement für Volkswirtschaft und Bildung des Kantons Wallis, 2018, S. 62f.) am Ende der 9. Klasse (11H) in beiden Fremdsprachen angestrebt wird (Peyer et al., 2016b, S. 47). Im Gegensatz dazu erreichen rund zwei Drittel der Lernenden im Englischunterricht die Lehrplanziele bereits in der 8. Klasse (10H) (Peyer et al., 2016b, S. 47). Die Unterschiede zwischen den Sprachkompetenzen in der ersten und zweiten Fremdsprache sind deutlich (Peyer et al., 2016b, S. 53). Es gilt jedoch zu

beachten, dass in der untersuchten BKZ-Region Englisch als erste und Französisch als zweite Fremdsprache unterrichtet werden (Peyer et al., 2016b, S. 50). Aufgrund der Ergebnisse gehen Peyer et al. (2016) davon aus, dass es unwahrscheinlich ist, dass sich die Kompetenzen im Fach Französisch bis zum Ende der obligatorischen Schulzeit den Kompetenzen im Fach Englisch angleichen werden und damit vergleichbar wären. In der vorliegenden Untersuchung wird im Gegensatz zur Studie von Peyer et al. (2016b) Französisch in der 5H als erste Fremdsprache eingeführt und Englisch ab der Klasse 7H als zweite Fremdsprache gelehrt.

Manno & Cuenat (2018, S. 219) schlagen vor, dass die Auswirkungen der Sprachenreihenfolge Französisch vor Englisch und der Mehrsprachigkeitsdidaktik auf die Wirksamkeit des Fremdsprachenunterrichts in der Passepartout-Sprachregion weiter untersucht werden und argumentieren (ebd., S. 237), dass allein die Festlegung von Unterrichtsstunden kein Garant dafür sein muss, dass das Ziel von HarmoS (vgl. Art. 4, HarmoS-Konkordat, 2008) erreicht werde. Weitere Faktoren, die in Bezug auf die beiden schulischen Fremdsprachen Französisch und Englisch in der Forschungsliteratur diskutiert werden, sind beispielsweise das Verbesserungspotenzial in der Anwendung der Prinzipien der Mehrsprachigkeitsdidaktik (z.B. Bono, M., 2008, S. 147ff.), die Förderung der Qualität des Fremdsprachenunterrichts durch die Fortbildung von Lehrpersonen (z.B. Manno & Cuenat, 2018, S. 237), der Altersfaktor beim Sprachenlernen (z.B. Lambelet & Berthele, 2015, S. 81f.), aber auch motivationale Aspekte (z.B. Heinzmann, 2013; Peyer et al., 2016a), welche mit dem schulischen Fremdspracherwerb einhergehen. Vor allem Ergebnisse aus der Forschung zur Motivation beim Sprachenlernen zeigen, dass Englisch sich einer grösseren Beliebtheit erfreut als Französisch (z.B. Stöckli, 2004; Heinzmann, 2013, Peyer et al., 2016a). Nicht nur die Beliebtheit als Schulfach, sondern auch der Status der englischen Sprache in Bezug auf verschiedene Faktoren, wie Wirtschaft und Globalisierung, lassen vermuten, dass die lexikalische Vielfalt in den englischsprachigen Textproduktionen höher ausfällt als in den französischsprachigen Aufsätzen.

Die vorliegenden Hypothesen haben das Ziel, die lexikalische Vielfalt nicht nur mit algebraischen Massen zu berechnen, sondern auch aus einer subjektiven Perspektive zu untersuchen. Hierbei sollen die Einschätzungen von Gutachterinnen und Gutachtern genutzt werden, um zu erforschen, wie zuverlässig subjektive Einschätzungen der lexikalischen Vielfalt sind und ob sie mit algebraischen Massen korrelieren. Zudem sollen zur Überprüfung der

Vergleichbarkeit zwischen den beiden Fremdsprachen verschiedene Zusammenhänge untersucht werden, wie etwa der Vergleich zu einem Referenzkorpus.

Basierend auf dem theoretischen Hintergrund und ausgehend von den Erwartungen des Lehrplans 21 in Bezug auf die Vergleichbarkeit der Kompetenzen im Bereich Schreiben in den beiden Fremdsprachen Französisch und Englisch am Ende der obligatorischen Schulzeit werden mit dieser Arbeit Aspekte der lexikalischen Vielfalt anhand folgender Hypothesen untersucht:

1. Textproduktionen von Lernenden am Ende der obligatorischen Schulzeit weisen in der zweiten Fremdsprache Englisch eine höhere algebraisch berechnete lexikalische Vielfalt aus als deren Textproduktionen in der ersten Fremdsprache Französisch.
2. Die lexikalische Vielfalt wird in den englischen Textproduktionen von menschlichen Ratern als höher eingestuft als in den französischen Textproduktionen.
3. Die algebraischen Masse der lexikalischen Vielfalt korrelieren in der ersten und zweiten Fremdsprache mit der Beurteilung durch menschliche Rater.
4. Die Wahrnehmung der lexikalischen Vielfalt korreliert in der ersten und in der zweiten Fremdsprache mit der Wahrnehmung der Erreichung der Aufgabenstellung, mit der Wahrnehmung bezüglich der Grammatik und der Rechtschreibung der Textproduktionen.
5. Die Wahrnehmung der lexikalischen Vielfalt, der Erreichung der Aufgabenstellung, der Grammatik und der Rechtschreibung korreliert mit dem Notendurchschnitt der Schülerinnen und Schüler am Ende der obligatorischen Schulzeit im Bereich Schreiben in der ersten und zweiten Fremdsprache.
6. Die Masse der lexikalischen Vielfalt weichen in den französischen Textproduktionen stärker von einem gesammelten Referenzkorpus ab als die englischen Textproduktionen.

7. Schülerinnen und Schüler, die in der Unterrichtssprache Deutsch eine hohe lexikalische Vielfalt vorweisen, haben auch in der ersten und zweiten Fremdsprache eine hohe lexikalische Vielfalt.

3.3 Design

Das vorliegende Forschungsdesign ist quasi-experimentell, korrelativ und deskriptiv. Die unabhängigen Variablen sind die Fremdsprachen, nämlich Englisch und Französisch. Die abhängige Variable ist die lexikalische Vielfalt, die auf verschiedene Weise gemessen wird, einschliesslich algebraischer Berechnungen und Bewertungen durch menschliche Rater, welche die Wahrnehmung der lexikalischen Vielfalt, die Erreichung der Aufgabenstellung, die Grammatik und die Rechtschreibung beinhalten.

Die Stichprobe besteht aus drei Konditionen: Die erste Kondition bildet die Untersuchungsgruppe und beinhaltet Schülerinnen und Schülern, die als erste Fremdsprache Französisch ab der Klasse 5H und als zweite Fremdsprache Englisch ab der Klasse 7H lernen, am Ende der obligatorischen Schulzeit. In der zweiten und dritten Kondition sind die Lernenden, deren Unterrichtssprache entweder Französisch oder Englisch ist. Die Datenerhebung erfolgt in der ersten Kondition durch schriftliche Textproduktionen der Teilnehmerinnen und Teilnehmer in Deutsch, Französisch und Englisch, deren lexikalische Vielfalt einerseits durch algebraische Masse (TTR, MTLD und Guiraud) berechnet wird und andererseits von angehenden Lehrpersonen in der Rolle von menschlichen Ratern bewertet wird. In der Vergleichsgruppe werden Aufsätze in der jeweiligen Unterrichtssprache gesammelt und anhand algebraischer Masse lexikalischer Vielfalt ausgewertet.

Die Daten werden mit Hilfe von Korrelations- und Regressionsanalysen sowie deskriptiven Statistiken untersucht, um die Hypothesen (vgl. Kapitel 3.2) zu überprüfen. Zusätzlich wird eine Analyse durchgeführt, um die Masse der lexikalischen Vielfalt in den Textproduktionen der Schülerinnen und Schüler mit einem auf derselben Aufgabe basierenden Referenzkorpus zu vergleichen.

Insgesamt zielt dieses Forschungsdesign darauf ab festzustellen, ob es signifikante Unterschiede oder Zusammenhänge zwischen der lexikalischen Vielfalt in den Textproduktionen von Französisch- und Englischlernenden am Ende der obligatorischen Schulzeit gibt und ob diese mit anderen Faktoren wie der Bewertung durch menschliche Rater oder Schulnoten korrelieren.

3.4 Methode

Polio (2012, S. 139) stellt in ihrer Arbeit neben der Fallanalyse acht Methoden (Umfragen, Interviews, Meta-Analysen, Beobachtung im Klassenzimmer, Ethnographie, Inhaltsanalyse, Textanalyse und Prozessforschung) zur spezifischen Untersuchung der Schriftlichkeit in der L2 und der L3 vor, wobei die verschiedenen Methoden auch überlappen können. Die gegenwärtige Arbeit ist auf die Textanalyse ausgerichtet. Die Daten, welche in der vorliegenden Arbeit ausgewertet werden, stammen aus den Erhebungen von den Schülerinnen und Schülern aus fünf Schulzentren im deutschsprachigen Oberwallis am Ende der obligatorischen Schulzeit. Im Bewusstsein, dass der Vergleich von Messungen in zwei Sprachen aufgrund der typologischen Unterschiede zwischen diesen Sprachen sehr schwierig ist, werden in beiden Sprachen jeweils zwei Klassen als Vergleichsgruppen herangezogen (vgl. Kapitel 2.4.2).

3.4.1 Textsammlung

Im Folgenden wird ein Überblick über die Datenerhebung gegeben, welche für die vorliegende Studie durchgeführt wurde. Es werden sowohl die Schreibaufgabe selbst, als auch der Ablauf der Pilotierungsphase und die verwendeten Befragungsinstrumente beschrieben. Zudem wird die Durchführung der Korrekturarbeiten und die Beschaffung eines Referenzkorpus erläutert.

3.4.1.1 Überblick

Im Rahmen dieser Forschungsarbeit werden schriftliche Textproduktionen in der Unterrichtssprache, sowie der ersten Fremdsprache Französisch und der zweiten Fremdsprache Englisch von Schülerinnen und Schülern am Ende der obligatorischen Schulzeit gesammelt. Vor der eigentlichen Datenerhebung wurde die Aufgabenstellung in drei Klassen pilotiert, um sicherzustellen, dass sie den Anforderungen entspricht und von den Teilnehmenden verstanden wird. Anschliessend wurde bei der zuständigen Dienststelle für Unterrichtswesen die Bewilligung eingeholt, um Klassen für die Studie anwerben zu dürfen. Die Schulleitungen der ausgewählten Schulen wurden über den Forschungsinhalt informiert und um ihre Bewilligung gebeten. Insgesamt konnten über die Schulleitungen und die Klassenlehrpersonen neun Klassen zur Teilnahme engagiert werden.

Die Datenerhebung fand Ende Mai und Anfang Juni 2022 statt, um sicherzustellen, dass die Schülerinnen und Schüler am Ende der obligatorischen Schulzeit waren. Die gesammelten Textproduktionen wurden anschliessend formal überarbeitet, damit sie mit Excel und R Studio analysiert werden konnten. Zusätzlich zu den Textproduktionen wurde ein Fragebogen zur Erhebung der sozio-demografischen Daten sowie der Einstellung gegenüber den jeweiligen Fremdsprachen, welche in der gegenwärtigen Untersuchung nicht berücksichtigt wird, eingesetzt. Die Klassenlehrpersonen stellten anonymisierte Daten zur Einstufung des Niveaus und zu den Notendurchschnitten der Schülerinnen und Schülern im Bereich Schreiben zur Verfügung. Um die Ergebnisse der Studie zuverlässiger interpretieren zu können, wurde zwischen Oktober 2022 und Januar 2023 ein Referenzkorpus erstellt, der aus Textproduktionen zur selben Schreibaufgabe von Schülerinnen und Schülern des jeweiligen Sprachgebiets besteht.

3.4.1.2 Schreibaufgabe

Die Entscheidung zur Aufgabenstellung für die Schreibaufgabe wurde sorgfältig getroffen, da audiovisuelle Medien vermutlich für Schülerinnen und Schüler im Schulalter attraktiver sind als Bildergeschichten. Aus diesem Grund wurde festgelegt, dass die Lernenden einen Trickfilm in Form eines kurzen Spots nacherzählen sollen. Ein wichtiges Kriterium bei der Auswahl der Filmreihe war, dass es sich um eine Serie handeln sollte, die aus mehreren, ungefähr gleich langen Episoden besteht, in denen die gleichen Protagonisten auftreten, damit ein Vergleich der Textproduktionen ermöglicht werden kann. Die Sequenzen sollten jedoch nicht zu lang sein, um eine schriftliche Darstellung des Inhalts zu erleichtern, aber dennoch ausreichend lang, damit die Handlung nicht zu kurz gefasst wird. Ein weiteres Kriterium bei der Filmauswahl war, dass die Filme entweder in den drei Sprachen Deutsch, Englisch und Französisch erhältlich sein müssen oder dass keine Sprache vorkommen darf. Besonderes Augenmerk wurde darauf gelegt, weil jede Schülerin und jeder Schüler zu jedem Spot einen Text in einer zufällig zugewiesenen Sprache verfassen muss.

Die Wahl fiel auf die kurzen Spots der Fernsehreihe «Die Sendung mit der Maus» des deutschen Medienverbunds ARD (vgl. Anhang 8.2. Videoclips). Die Audiowiedergabe der Spots besteht aus fiktiven Wortketten in Form von Tiergeräuschen, Umgebungsgeräuschen oder Toneinlagen, die universell verständlich sind. Die Länge der Episoden erscheint ideal, da sie zwischen einigen Sekunden und wenigen Minuten dauern und Geschichten mit einer

klassischen Erzählstruktur erzählen. Es wurde darauf geachtet, dass bei der Auswahl der Werbespots immer dieselben drei Charaktere auftreten, nämlich die Maus, der Elefant und die Ente. Jeder Spot präsentiert ein Ereignis mit diesen drei Protagonisten und alle Geschichten haben ein gutes Ende.

Die Aufgabenstellung für die Lernenden sollte klar strukturiert und ausformuliert sein. Die Formulierung der Aufgabenstellung spielt eine wichtige Rolle bei handlungsorientierten Testaufgaben, da die Lernenden eine definierte Schreibabsicht innerhalb eines Handlungskontextes erfüllen sollen. Es ist von Bedeutung, dass die Schreibkompetenz unabhängig von der Leseverstehenskompetenz der Lernenden getestet wird. Die Schreibaufgaben sind deshalb auch für die Instruktionen in den Fremdsprachen in der Unterrichtssprache verfasst und nicht in der Zielsprache. Durch die Verwendung der Unterrichtssprache bei der Formulierung der Aufgabenstellung wird sichergestellt, dass möglichst viele Schülerinnen und Schüler die Aufgabenstellung innerhalb kürzester Zeit verstehen können (vgl. Lenz, Peter et al., 2009, S. 48).

In der Pilotphase im Mai 2021 werden insgesamt drei Klassen der Stufen 9H (N=14) und 11H (N=19) damit beauftragt, den Videoclip für einen Austauschschüler aus dem französischsprachigen Unterwallis in einem kurzen Text auf Französisch zusammenzufassen und dabei möglichst detailliert auf die einzelnen Handlungen einzugehen und zu beschreiben, was sie in der Episode von «Der Sendung mit der Maus» gesehen haben (vgl. Kapitel 8.1.1). Die Schülerinnen und Schüler wurden schriftlich mit der Aufgabenstellung vertraut gemacht. Sie sahen den Spot einmal und konnten sich dabei Notizen machen. Es war ihnen nicht gestattet einen Diktionär oder ein Online-Wörterbuch oder gar einen Online-Übersetzer zu verwenden. Es wurde darauf hingewiesen, dass nur die Zielsprache verwendet werden darf und ein Wort, falls es nicht bekannt war, aber für die Textproduktion benötigt wurde, ebenfalls in der Zielsprache umschrieben werden sollte. Die Lernenden verfassten ihre Aufsätze handschriftlich in einer sauberen Darstellung und einer leserlichen Schrift im vorgegebenen Raster. Das Zeitfenster war auf dreissig Minuten beschränkt.

Folgende Anpassungen wurden nach der Auswertung der Daten aus der Pilotierung vorgenommen (vgl. Kapitel 8.1.2):

- Das Entziffern der Handschrift und das Transkribieren der Textproduktionen stellte sich als grosse und zeitaufwändige Herausforderung heraus, weshalb in der vorliegenden Untersuchung die Textproduktionen via FORMS eingeholt werden, damit diese direkt digital vorliegen und entsprechend weiterverarbeitet werden können.
- Die Schülerinnen und Schüler geben mittels Selbständigkeitserklärung an, dass bei der Bewältigung der Aufgabe keine Hilfsmittel (z.B. Übersetzer, Wörterbücher, etc.) verwendet werden und die Aufgabe selbständig gelöst wird. Dabei handelt es sich in FORMS um eine Pflichtfrage, d.h. dass die Lernenden keinen Zugang zum Videoclip und Textfeld haben, wenn die Selbständigkeitserklärung nicht bestätigt wird.
- Das Zeitfenster von 30 Minuten wurde in der Pilotierungsphase nicht genutzt, weshalb der zeitliche Rahmen für das Verfassen des Textes bei der Durchführung auf 20 Minuten gekürzt wurde.
- Damit die Textproduktionen in den verschiedenen Sprachen eines Schülers oder einer Schülerin einander zugeordnet werden können, erhält jeder Schüler und jede Schülerin einen Textcode zugeteilt, welcher bei allen Textproduktionen sowie bei der Ermittlung weiterer Daten mittels Fragebogen oder mittels Umfrage bei den Klassenlehrpersonen angegeben werden muss.
- Die Aufgabenstellung wird für die Erhebung der Texte in den anderen Sprachen (Deutsch und Englisch) so adaptiert, dass der Inhalt des Clips in der deutschen Textproduktion für eine kranke Mitschülerin aufgeschrieben wird, währenddem sich der englische Text an eine Schülerin aus Amerika richtet.
- Der Videoinput aus der Pilotphase wurde mit zwei weiteren Videospots zu «Der Sendung mit der Maus» ergänzt (vgl. Anhang 8.2. Videoclips), sodass ein komplettes Counterbalancing für ein within-subjects-Experiment mit drei Stimulussets möglich ist (vgl. Tabelle1).

Tabelle 1: Counterbalancing bei der Datenerhebung

	Schüler 1	Schüler 2	Schüler 3	Schüler 4	Schüler 5	Schüler 6	Schüler 7	Schüler 8	...
A	DE	DE	FR	FR	EN	EN	DE	DE	...
B	FR	EN	DE	EN	DE	FR	FR	EN	...
C	EN	FR	EN	DE	FR	DE	EN	FR	...

(A: Videoinput 1, B: Videoinput 2, C: Videoinput 3, DE = Deutsch, FR = Französisch, EN = Englisch)

Auch diese Inputs wurden unter Berücksichtigung der erwähnten Aspekte von zwei Klassen der 9H (N=31) pilotiert, um einerseits den zeitlichen Rahmen zu bestätigen und andererseits Textmaterial auch in deutscher und englischer Sprache zu erhalten, welches als Beispiel bei der Instruktion der Gutachterinnen und Gutachter eingesetzt werden konnte.

3.4.1.3 Erhebungen

Nach Erhalt der Zustimmung der Dienststelle für Unterrichtswesen und Einwilligung der Schulleitungen, wurden die Klassenlehrpersonen der entsprechenden 11H Klassen angeworben. Die Lehrpersonen wurden direkt kontaktiert und persönlich instruiert, um sicherzustellen, dass sie in der Lage sind, die Erhebung eigenständig vor Ort zu organisieren und nur bei Bedarf eine direkte Intervention erforderlich ist. Zu diesem Zweck wurde eine detaillierte Anleitung für die Lehrpersonen erstellt und persönlich besprochen. Schliesslich nehmen insgesamt sieben Klassen aus fünf Schulzentren an der Datenerhebung teil, obwohl ursprünglich neun Klassen aus sechs Schulzentren zur Verfügung standen.

Die Lernenden der Untersuchungsgruppe haben im Mai und Juni 2022 jeweils einen online-Fragebogen zur Erhebung der sprachbiografischen Angaben eingereicht. In diesem Fragebogen haben die Schülerinnen und Schüler u.a. geschlossene Fragen zu biografischen und demografischen Informationen, wie etwa Geschlecht, Alter oder Erstsprache beantwortet. Ausserdem wurden mit dem Fragebogen, basierend auf bereits existierenden Testinstrumenten (Stöckli, 2004; Gardner Robert C., 2004; Dörnyei & Ushioda, 2011; Dörnyei & Chan, 2013; Peyer et al., 2016b), Daten zur Beliebtheit der Sprachenfächer, zur Selbsteinschätzung in Bezug auf die schriftlichen Kompetenzen in den verschiedenen Sprachen, zur Einstellung und zur Motivation zum jeweiligen Fremdsprachenunterricht und zur Fremdsprache erhoben, welche aber in der vorliegenden Arbeit nicht weiter analysiert werden (vgl. Kapitel 8.2). Anschliessend haben die Lernenden im selben Zeitraum an drei voneinander unabhängigen Terminen die Textproduktionen vorgenommen. Mittels Formular wurden die Lehrpersonen nach den Niveauezuteilungen und den Notendurchschnitten der Lernenden im aktuellen Jahr in den jeweiligen Sprachen im Bereich Schreiben befragt (vgl. Kapitel 8.3). Es ist hier zu beachten, dass die Unterrichtssprache Deutsch und die Fremdsprache Französisch als Niveaufächer mit den Abstufungen Niveau I und Niveau II

unterrichtet werden, während dem der Englischunterricht in ganzheitlich heterogenen Gruppen stattfindet.

Tabelle 2: Textproduktionen nach Sprache und Modus

	DE	FR	EN	Total:
A	34	35	38	107
B	33	33	32	98
C	34	35	32	101
Total:	101	103	102	306

(A: Videoinput 1, B: Videoinput 2, C: Videoinput 3, DE = Deutsch, FR = Französisch, EN = Englisch)

Die Erhebungen der Textproduktionen in den Vergleichsgruppen fand zwischen Oktober 2022 und Januar 2023 statt, wobei sich die Suche nach einer geeigneten Gruppe v.a. für die Sprache Englisch als sehr schwierig gestaltete. Insgesamt wurden 23 Schulen in Grossbritannien und den USA zufällig per E-Mail ausgewählt und kontaktiert. Nur fünf dieser Schulen haben geantwortet, jedoch hat keine Schule ihre Bereitschaft signalisiert, die Textproduktion im Schulzentrum durchzuführen. Es ist anzumerken, dass die Vergleichsgruppe lediglich eine Textproduktion durchführt. Leider konnte aufgrund von Zeitmangel oder der fehlenden Zustimmung der Schulleitung auch keine Erhebung in Klassen durchgeführt werden, welche von der Universität Freiburg vermittelt wurden. Letztendlich haben sich jedoch zwei über Dritte vermittelte Lehrpersonen aus London (Grossbritannien) und Oregon (USA) bereit erklärt, an der Untersuchung teilzunehmen.

Die nachstehende Tabelle 3 veranschaulicht die Anzahl der Textproduktionen der beiden Vergleichsgruppen.

Tabelle 3: Anzahl Textproduktionen nach Vergleichsgruppe und Modus

	FR	EN
A	13	16
B	12	13
C	12	14
Total:	37	43

(A: Videoinput 1, B: Videoinput 2, C: Videoinput 3, FR = Französisch, EN = Englisch)

3.4.1.4 Lernende

Die vorliegende Studie bezieht sich auf eine Untersuchungsgruppe von Schülerinnen und Schülern, welche die obligatorische Schule in der Unterrichtssprache Deutsch im deutschsprachigen Wallis absolvieren. Zudem werden Daten in zwei Vergleichsgruppen erhoben, wobei die Unterrichtssprache in einer Gruppe Französisch und in der anderen Gruppe Englisch ist. Eine Übersicht über die soziodemografischen Merkmale der Lernenden wird in Tabelle 4 präsentiert.

Tabelle 4: Soziodemografische Daten (N=183)

		Jahrgang	Geschlecht	Muttersprache
Untersuchungs- gruppe Datenerhebung im Mai und Juni 2022	Unterrichtssprache Deutsch			Deutsch: 69
				Französisch: 11
		2004: -		Englisch: -
	7 Klassen aus 5	2005: 1		Portugiesisch: 12
	Schulzentren	2006: 84	m: 39	Albanisch: 3
	(N = 103)	2007: 16	w: 62	Bosnisch: 1
		2008: -	NA: 2	Italienisch: 1
		2009: -		Andere: 3
		NA: 2		NA: 2
Vergleichsgruppe Datenerhebung zwischen Oktober 2022 und Januar 2023	Unterrichtssprache Französisch			Deutsch: -
				Französisch: 34
		2004: 1		Englisch: -
	2 Klassen aus dem	2005: -	m: 20	Portugiesisch: 1
	gleichen	2006: 1	w: 17	Albanisch: 1
	Schulzentrum	2007: 11		Italienisch: -
		2008: 30		Andere: 1
	(N= 37)	2009: -		
	Unterrichtssprache Englisch			Deutsch: -
				Französisch: 1
		2004: -		Englisch: 32
	2 Klassen aus	2005: -	m: 13	Portugiesisch: -
	verschiedenen	2006: -	w: 30	Albanisch: -
	Schulzentren	2007: 9		Italienisch: -
		2008: 19		Andere: 10
	(N=43)	2009: 9		

Die vorliegenden soziodemografischen Daten (vgl. Tabelle 4) zeigen, dass die Untersuchungsgruppe nicht ausschliesslich aus monolingualen Schülerinnen und Schülern besteht, sondern auch anderssprachige Kinder umfasst. Diese Untersuchung ist nicht darauf ausgerichtet, Schlussfolgerungen über die Muttersprache zu ziehen, weshalb der Fokus auf

den unterrichteten Sprachen Deutsch als Unterrichtssprache, der ersten Fremdsprache Französisch und der zweiten Fremdsprache Englisch liegt. Insgesamt nehmen 103 Schülerinnen und Schüler an der Datenerhebung der Untersuchungsgruppe teil, wobei es bei zwei Lernenden keine persönlichen Angaben gibt, da diese zum Zeitpunkt der Erhebung der soziodemografischen Daten abwesend waren.

Zum Zeitpunkt der Datenerhebung sind die Lernenden der Untersuchungsgruppe zwischen 15 und 17 Jahre alt, wobei die Mehrheit (n=62) 16 Jahre alt ist. Insgesamt nahmen 62 Schülerinnen und 39 Schüler an der Datenerhebung in der Untersuchungsgruppe teil. Die Mehrheit der Schülerinnen und Schüler der Untersuchungsgruppe gibt an, Deutsch als Muttersprache zu haben, wobei Muttersprache in diesem Kontext als die Sprache interpretiert wird, die eine Person als primäre Sprache in ihrer Kindheit erworben hat und die Identität und kulturelle Zugehörigkeit ausdrückt (vgl. Davies, 2003, S. 207). Die zweithäufigste Muttersprache in der Untersuchungsgruppe war Portugiesisch. Es gab auch elf Schülerinnen und Schüler mit französischer Muttersprache, die das letzte Jahr der obligatorischen Schulzeit in deutscher Sprache im Rahmen eines Immersionsjahres absolvieren. Die Textproduktionen dieser Schülerinnen und Schüler werden in der Auswertung nicht berücksichtigt.

3.4.1.5 Korrekturarbeit

Da die Erhebung der Textproduktionen der Untersuchungsgruppe an jeweils verschiedenen Terminen stattfand, sind nicht in allen Sprachen gleich viele Texte entsprechend der Anzahl Teilnehmenden (n=103) eingegangen (vgl. Kapitel 3.4.1.3. Erhebungen). In Tabelle 5 ist ersichtlich, dass die Anzahl der Texte, die schlussendlich in der Analyse berücksichtigt werden, nicht der Anzahl der eigentlichen Textproduktionen entspricht (vgl. Tabelle 2). Dies hat verschiedene Gründe: Einerseits werden die Texte der französischsprachigen Immersionsschülerinnen und Immersionsschüler nicht berücksichtigt. Andererseits können in den französischsprachigen Textproduktionen drei Texte aufgrund der Textlänge von eins bis acht Wörtern nicht berücksichtigt werden, um zu vermeiden, dass diese das Ergebnis verzerren. Dabei handelt es sich um die Texte «vert, elefant petit, tomer, mettre», «elephant» oder «Le rat et une elefant kochen a mais». In derartigen Situationen ist es sowohl aus mathematischer als auch aus konzeptioneller Perspektive nicht angemessen, die lexikalische Vielfalt sowie die Korpusfrequenz zu quantifizieren, da diese Texte aus einzelnen Wörtern

oder Satzfragmenten bestehen. Die französischen Textproduktionen dreier weiterer Personen müssen ausgeschlossen werden: Eine Person hat die Aufgabe mit «je ne komprade pas» gelöst, während eine weitere Person diesen Text «The mouse wanted to go over the street but I don't know more because I didn't watch the film I am sorry» einreicht. Die dritte Person hat den Text auf Portugiesisch statt auf Französisch geschrieben.

Tabelle 5: Analyzierte Textproduktionen nach Sprache, Modus und Gruppe

	Französisch		Englisch		Deutsch
	UG	VG	UG	VG	UG
A	28	13	30	16	33
B	30	12	29	13	28
C	29	12	30	14	29
<i>Total:</i>	88	37	89	43	90

(A: Videoinput 1, B: Videoinput 2, C: Videoinput 3, UG = Untersuchungsgruppe, VG = Vergleichsgruppe)

Alle Texte werden im Anschluss bearbeitet, um schliesslich vier unterschiedliche Versionen zu erhalten, die bei der Berechnung der algebraischen Masse der lexikalischen Vielfalt verglichen werden können. Diese Versionen beinhalten eine unkorrigierte Fassung, eine zweite Version, in der u.a. orthografische Fehler korrigiert werden, eine Version, in der Fehler durch Ersetzen behoben werden und eine Version, in der Fehler durch Löschung eliminiert werden, was in folgender Tabelle 6 aufgezeigt wird.

Tabelle 6: Probleme und Massnahmen bei der Textbearbeitung

Problem:	Massnahme:			
	Version 1 TextOrig	Version 2 TextKorr	Version 3 TextXXX	Version 4 TextDelete
Umgang mit Fehlern in der Anwendung der Rechtschreibung z.B. <i>withe</i> -> <i>white</i> , <i>fance</i> -> <i>fence</i> , <i>elefant</i> -> <i>éléphant</i> , <i>etc.</i>	KK*	Korrektur, sofern das Wort aus dem Zusammenhang oder durch lautgetreues Schreiben erkennbar.		
Umgang mit Gross- und Kleinschreibung	KK*	KK*	KK*	KK*

Wörter oder Satzfragmente stehen in der L1 <i>z.B. andre ceite, Maus, Sack, Topf, ente, pfanne, genug, exploriert etc.</i>	KK*	Taggern	Ersetzen durch XXX	Löschen aus dem Text
Umgang mit nicht erkennbaren Wörtern <i>z.B. room in the circh, explochen, gepisen, tven, moiser, sleth, etc.</i>	KK*	Taggern	Ersetzen durch XXX	Löschen aus dem Text
Umgang mit Lauten, erfundenen oder undefinierbaren <i>z.B. bruhhhh, quaq quaq, buuuuuuum, BOOM, etc.</i>	KK *	Taggern	Ersetzen durch XXX	Löschen aus dem Text
Umgang mit arabischen Zahlen	KK*	Taggern	Taggern	Taggern

*KK = Keine Korrektur

Die Verwendung von Lemmaeinheiten erweist sich als besonders praktisch, da Lernende häufig Rechtschreib- und Grammatikfehler aufweisen, die sich negativ auf die lexikalische Analyse auswirken können. Zudem weisen Treffers-Daller et al. (2016, S. 11) darauf hin, dass die Lemmatisierung unerlässlich ist, wenn die lexikalische Vielfalt in typologisch unterschiedlichen Sprachen vergleichbarer gemacht werden sollen (vgl. Kapitel 2.4.2). Aus diesem Grund werden die vier Versionen einer automatischen Lemmatisierung mittels des R-Pakets UDPipe unterzogen.

3.4.2 Ratings

Im Anschluss an die Textauswahl wurden unabhängige Gutachterinnen und Gutachter beauftragt, die Texte anhand von vier Kriterien zu bewerten. Das Rating erfolgte holistisch und erforderte von den Gutachterinnen und Gutachtern eine einzige globale Bewertung jedes Textes auf einer 5-Punkte Likert-Skala (vgl. Jarvis, 2013b, 2017; Jarvis & Daller, 2013; Bonvin & Lambelet, 2017, S. 69; Vanhove et al., 2019, S. 502ff.; Konsortium ÜGK, 2019, S. 49). Eine deskriptive Beschreibung des zu beurteilenden Textkorpus, die verschiedenen Bewertungskriterien, die von den Gutachterinnen und Gutachtern benutzt wurden, sowie ein detaillierter Überblick über den Ratingablauf werden in den kommenden Abschnitten beschrieben.

3.4.2.1 Texte

Bei den Textproduktionen der Schülerinnen und Schüler, welche den Gutachterinnen und Gutachtern vorgelegt wurden, handelt es sich um die unkorrigierte Originalversion der

Lernenden (vgl. Tabelle 5). Die folgende Tabelle (vgl. Tabelle 7) gibt einen Überblick über die Anzahl der Token in den jeweiligen Texten nach der Lemmatisierung mit UDPipe.

*Tabelle 7: Deskriptive Analyse des Textkorpus in Bezug auf Tokens**

	Untersuchungsgruppe			Vergleichsgruppe	
	Französisch (N=88)	Englisch (N=89)	Deutsch (N=90)	Französisch (N=37)	Englisch (N=43)
Min. Anzahl Tokens	8	15	34	33	8
1. Quartil (25%)	40.75	61	79.5	75	39
Median	69	109	129	106	78
Mittelwert	72	117	139	105.24	104.05
3. Quartil (75%)	92	154	181.25	125	117
Max. Anzahl Tokens	223	380	362	194	565

*Ntok2_TextOrig (nach der Auflösung von Kontraktionen und der Eliminierung von Zahlen)

Ein Vergleich der Sprachen in der Untersuchungsgruppe anhand der Tabelle 7 zeigt, dass die Anzahl der Token in den Texten im Durchschnitt am höchsten bei den deutschen Texten ist (Mittelwert = 139), gefolgt von der Anzahl in den englischen (Mittelwert = 117) und den französischen Textproduktionen (Mittelwert = 72). Die Anzahl der Token variiert auch innerhalb jeder Sprache erheblich, wie durch den Unterschied zwischen dem Minimum und Maximum dargestellt wird. Ein Vergleich der Quartile zeigt, dass die deutschsprachigen Texte tendenziell länger sind als die der anderen beiden Gruppen. Der Median der deutschen Aufsätze (129 Token) ist auch höher als derjenige der englischen (109 Token) und der französischen Texte (69 Token).

Wenn die Französischtexte der Untersuchungsgruppe mit denjenigen der Vergleichsgruppe verglichen werden, zeigt die Tabelle 7, dass die Texte in der Untersuchungsgruppe tendenziell kürzer sind als in der Vergleichsgruppe. Der Median der französischen Aufsätze in der Untersuchungsgruppe beträgt 69 Token, während der Median in der Vergleichsgruppe 106 Token beträgt. Der Mittelwert für die Anzahl der Wörter bei den französischen Textproduktionen der Untersuchungsgruppe (72 Token) ist ebenfalls niedriger als der Mittelwert in der Vergleichsgruppe (105,24 Token).

Bei den englischen Texten ist das Gegenteil der Fall. Werden die englischen Textproduktionen der beiden Gruppen gegenübergestellt, lässt sich zeigen, dass die Texte der Untersuchungsgruppe tendenziell länger sind als diejenigen der Vergleichsgruppe. Der Median der Englischtexte in der Untersuchungsgruppe beträgt 109 Token, im Vergleich zum Median der Vergleichsgruppe von 78 Token. Dies widerspiegeln ebenfalls die Mittelwerte der Anzahl Token, mit 117 Token in der Untersuchungsgruppe und 104.05 Token in der Vergleichsgruppe.

Insgesamt lässt sich mit Tabelle 7 aufzeigen, dass die Anzahl der Token in den Texten erheblich variiert, sowohl innerhalb jeder Sprache als auch zwischen den Sprachen. Deutschsprachige Texte in der Untersuchungsgruppe sind im Durchschnitt länger als die englischen und französischen Texte. Die französischen Texte der Untersuchungsgruppe sind im Vergleich zu denen der Vergleichsgruppe tendenziell kürzer, während die englischen Texte der Untersuchungsgruppe im Vergleich zu denen der Vergleichsgruppe tendenziell länger sind.

3.4.2.2 Raters

Die Raters, welche die Texte der Schülerinnen und Schüler in Bezug auf die Aufgabenstellung, die Grammatik, die Rechtschreibung und die lexikalische Vielfalt bewerten, sind angehende Lehrpersonen, welche zum Zeitpunkt der Ratings im Abschlussjahr der pädagogischen Hochschule Wallis (PH-VS) im Studiengang Bachelor Primarstufe studieren.

Tabelle 8: Daten zum Rater Sample (N=18)

Alter	Geschlecht	Sprachkenntnisse		
		Deutsch	Französisch	Englisch
Ø 25.9 Jahre	m: 8 w: 10	Muttersprache: 18	B1: 3	B1: 1
			B2: 13	B2: 11
			C1: 2	C1: 6

Alle Studierenden haben mindestens einen Maturaabschluss, zwei haben bereits einen Bachelor absolviert und eine Person hat die höhere Fachschule absolviert. Die zukünftigen Lehrpersonen sind zwischen 22 und 37 Jahre alt. Alle Studierenden sind deutscher

Muttersprache, wobei eine Person aus Deutschland stammt und die restlichen Raters aus der Schweiz kommen. Gemäss den internen Weisungen der PH-VS (vgl. Art. 4, Departement für Bildung und Sicherheit, 2016), müssen die Studierenden bis zum Ende des fünften Semesters ein Niveau B2 nach dem GER in den beiden Fremdsprachen Französisch und Englisch vorweisen können, damit das Studium ordnungsgemäss abgeschlossen werden kann. Die Mehrheit der Studierenden kann zum Zeitpunkt der Datenerhebung bereits ein Niveau B2 oder C1 nach dem GER ausweisen.

3.4.2.3 Pilotierung

Ausgehend vom GER und dem Lehrplan 21 (vgl. Kapitel 2.2) werden in Anlehnung an bisherige Studien (z.B. Jarvis, 2013b; Treffers-Daller, 2013; Jarvis & Daller, 2013; Jarvis, 2017; Bonvin & Lambelet, 2019; Vanhove et al., 2019, S. 507f.) zwei mögliche Szenarien mit Beurteilungen auf einer 5-Punkte Likert-Skala mit jeweils drei Studierenden ausserhalb der Ratinggruppe getestet. Das Rating wird auf einer 5-stufigen Likert-Skala vorgenommen, damit nicht auf das gängige Notensystem der Schule (Note 1 – 6) zurückgegriffen wird. Die Likert-Skala ist folgendermassen indexiert: 1= sehr schlecht, 2= eher schlecht, 3= mittelmässig, 4= eher gut, 5= sehr gut.

Das erste Szenario sieht vor, dass mittels holistischem Rating Daten zu vier verschiedenen Fragestellungen eingeholt werden, wobei die Fragestellungen mit Indikatoren präzisiert werden, die sich an die Formulierungen des GER und des Lehrplans 21 (vgl. Kapitel 2.1 und 2.2) anlehnen. Dieses Szenario sieht folgendermassen aus:

1. **Aufgabe:** Wie beurteilen Sie diesen Text in Bezug auf das Erfüllen der Aufgabenstellung und den Inhalt?
 - Das Ereignis wird ausführlich und detailliert beschrieben.
 - Eine Beschreibung sämtlicher Inhalte ist gegeben.
 - Die Inhalte werden sachlich richtig zusammengefasst.
 - Der Textinhalt ist nachvollziehbar und Zusammenhänge sind erkennbar.
 - Die Sätze sind verständlich formuliert und passen zum Inhalt.

2. **Grammatik:** Wie beurteilen Sie diesen Text in Bezug auf die einfachen grammatischen Strukturen?
 - Es ist ein klarer und variantenreicher Satzbau vorhanden.
 - Die Grammatikregeln werden umgesetzt (z.B. Die häufige Konjugationsformen von Verben sind korrekt, etc.)
 - Der Text ist sprachlich korrekt.
3. **Rechtschreibung:** Wie beurteilen Sie diesen Text in Bezug auf die Rechtschreibung?
 - Die Rechtschreibregeln werden mehrheitlich umgesetzt.
 - Es kommen kaum störende Rechtschreibfehler vor.
 - Es kommen kaum unpassende und fehlerhafte Sätze vor.
4. **Lexikalische Vielfalt:** Wie beurteilen Sie diesen Text in Bezug auf die lexikalische Vielfalt?
 - Der Wortschatz, der verwendet wird, ist passend.
 - Der Wortschatz ist abwechslungs- und variantenreich.
 - Passende Satzanfänge und Satzverbindungen sind vorhanden.

In einem zweiten Szenario werden in einem holistischen Rating vier Aspekte mit Hilfe der Frage «Inwiefern stimmen Sie folgenden Aussagen zu?» abgefragt, wobei die Begutachterinnen und Begutachter ihre Einschätzung auf der gleichen fünfstufigen Likert-Skala abgeben. Im genannten Szenario werden folgende Aspekte beurteilt:

1. **Aufgabe:** Im Text wird mit einfachen Sätzen und Ausdrücken über vertraute Handlungen berichtet, wobei mit einfachen Gestaltungsmitteln eine ästhetische Wirkung erzielt wird.
2. **Grammatik:** Der Text beinhaltet einfache grammatische Strukturen.
3. **Rechtschreibung:** Der Text ist angemessen und korrekt geschrieben.
4. **Lexikalische Vielfalt:** Der Text beinhaltet ein genügend grosses Repertoire an Wörtern und Wendungen.

Für die Pilotierung des Ratings wurden französische und englische Textproduktionen aus den Pilotphasen (vgl. Kapitel 3.4.1.2) verwendet. Die Probanden in beiden Szenarien erhielten jeweils ein Aufgabenset von zwanzig zufällig gewählten Texten in französischer oder englischer Sprache und gaben das Rating mittels Textcode und anhand des zugewiesenen Szenarios über FORMS ab. Für jede Sprache wurden dieselben Skalen verwendet. Dadurch werden insgesamt vier holistische Skalen im Rating eingesetzt, um eine vergleichbare Beurteilung auch zwischen den Sprachen zu gewährleisten. Im anschliessenden Austausch über die beiden Szenarien hat sich gezeigt, dass ein Rating mit Indikatoren nicht nur zeitaufwändiger, sondern auch schwerfälliger ist und die Texte der Schülerinnen und Schüler auf die Indikatoren reduziert werden, wobei eine globale Sicht über die Aspekte kaum noch wahrgenommen werden kann. Zudem wird wie auch bei Vanhove et al. (2019, S. 507) darauf hingewiesen, dass die Gutachterinnen und Gutachter vorgängig exemplarisch Schülertexte sichten, damit sie sich einerseits mit der Aufgabenstellung vertraut machen und präzisere Einschätzung gemacht werden können.

Die Kriterien für das Rating werden deshalb folgendermassen angepasst:

Wie beurteilen Sie den Text in Bezug auf...?

1. ...das Erfüllen der Aufgabenstellung?
2. ...die grammatischen Strukturen?
3. ...die Rechtschreibung?
4. ...die lexikalische Vielfalt?

3.4.2.4 Ablauf

Die Studierenden, welche die Texte beurteilten, füllten zunächst einen Fragebogen auf Forms aus, um die soziodemografischen Daten der Gutachterinnen und Gutachter zu erheben (vgl. Tabelle 8). Die Gutachterinnen und Gutachter sahen sich ebenfalls die drei Videoclips an und konsultierten den Auftrag an die Schülerinnen und Schüler, damit ein Verständnis für den Inhalt des Textes vorliegt und die den Texten zugrunde liegende Aufgabenstellung klar war. D.h. die Bewertenden waren sich bewusst, dass die Texte von Lernenden der obligatorischen Schulzeit verfasst werden.

Damit das Rating erfolgreich ist und die Rückmeldungen aus der Pilotierung berücksichtigt werden konnten, diskutierten die Studierenden jeweils fünf ausgewählte Texte pro Sprache aus den Pilotierungen (vgl. Kapitel 3.4.1.2). Das Ziel dabei war es, sich mit dem Vorgehen und dem Beurteilen der Textproduktionen anhand der Fragen zur Einschätzung vertraut zu machen.

Anschliessend wurden die Studierenden gebeten, die mithilfe von Excel zufällig zugeteilten 79 bis 84 Texte (d.h. 24 bis 27 Texte pro Sprache) zu lesen und das Erfüllen der Aufgabenstellung, sowie die Wahrnehmung in Bezug auf die grammatischen Strukturen, die Rechtschreibung und die lexikalische Vielfalt auf der 5-Punkte-Skala zu bewerten, wobei die Werte gekennzeichnet sind (1= sehr schlecht, 2= eher schlecht, 3= mittelmässig, 4= eher gut, 5= sehr gut). Die Studierenden bewerteten zuerst alle Texte einer Sprache, bevor in einem zweiten und dritten Ratingdurchlauf die Texte der anderen Sprachen begutachtet wurden. Die Reihenfolge der Sprachen, sowie die Reihenfolge der Texte innerhalb der Sprachen war ebenfalls randomisiert, damit Gewöhnungs- und Koexistenz-Effekte vermieden werden konnten, wobei für die Gutachterinnen und Gutachter klar vorgegeben war, in welcher Reihenfolge die Ratings abgegeben werden.

Zudem wurden die Studierenden darauf hingewiesen, dass es nicht ungewöhnlich ist, wenn diese Aufgabe schwierig erscheint, dass jedoch die intuitive Wahrnehmung im Forschungsprojekt von Interesse ist. Deshalb sollten die Gutachterinnen und Gutachter nicht zu sehr versuchen nachzudenken, aber dennoch konsistent bei der Einschätzung sein (vgl. Vanhove et al., 2019, S. 507).

Da das Interesse an der Wahrnehmung der lexikalischen Vielfalt liegt, wird dieser Term in der Einführung der Ratings im Austausch über die Übungstexte mit den Studierenden thematisiert. Gemäss Jarvis (2013a, S. 52) wäre es ein Fehler, Gutachterinnen und Gutachter dahingehend zu schulen, wie sie die lexikalische Vielfalt von Texten beurteilen sollen, wenn das Ziel darin besteht, herauszufinden, ob Menschen bereits ein Verständnis von lexikalischer Vielfalt haben, welche Dimensionen der lexikalischen Vielfalt sie erkennen und wie diese in Beziehung zueinander stehen. Jarvis (2013a, S. 25) definiert die lexikalische Vielfalt als «the variety of word use that can be found in a person's speech or writing». Da es in der vorliegenden Arbeit auch darum geht, die lexikalische Vielfalt zu beurteilen, haben die Studierenden zwar keine einheitliche Definition von lexikalischer Vielfalt, wie diejenige von

Jarivs (2013a, S. 25) erhalten, sie haben sich aber mit dem Begriff der lexikalischen Vielfalt, sowie der Aufgabenstellung, der Grammatik und der Rechtschreibung, in Form eines Austausches untereinander anhand der fünf pilotierten Texte pro Sprache auseinandergesetzt.

Jeder Text wurde vier bis sechs Mal beurteilt, wobei jede Textproduktion durchschnittlich 5.34 Ratings erhält. Durch diese mehrfache Beurteilung von Texten durch verschiedene Gutachterinnen und Gutachter kann die Interrater-Reliabilität abgeschätzt werden, welche sich aus der internen Konsistenz der einzelnen Rater und der Strenge der Bewertung zusammensetzt.

3.5 Ergebnisse

Nach Abwägen verschiedener geeigneter Massen zur Auswertung der lexikalischen Vielfalt werden schliesslich die TTR, der Guiraud-Index und die MTLD berechnet (vgl. Kapitel 2.4.2). Der Guiraud-Index stellt einen der besten Prädiktoren für menschliche Ratings der lexikalischen Vielfalt dar (Vanhove et al., 2019, S. 520) und eignet sich für die Analyse kurzer Textproduktionen (Bonvin et al., 2018, S. 137). Auch die MTLD bietet sich als Mass für die Berechnung der lexikalischen Vielfalt von nahezu allen Texten an, obwohl sie bei kurzen Texten nicht empfohlen wird (Bonvin et al., 2018, S. 137). Da viele der gesammelten Textproduktionen mit einem Durchschnitt von 72 bis 139 Token eher kurz sind (vgl. Tabelle 7), machen die algebraischen Masse, bei denen Texte in mehrere Teile zerteilt werden, wie dies bei manchen anderen Massen der Fall ist (vgl. Kapitel 2.4.2), wie beispielsweise die Berechnung der MSTTR, die Berechnung von D oder HD-D, wenig Sinn. Die Literatur zeigt, dass bislang noch kein Mass existiert, das ohne Einschränkungen auch auf kurze Texte empfohlen werden kann (Bonvin et al., 2018, S. 137).

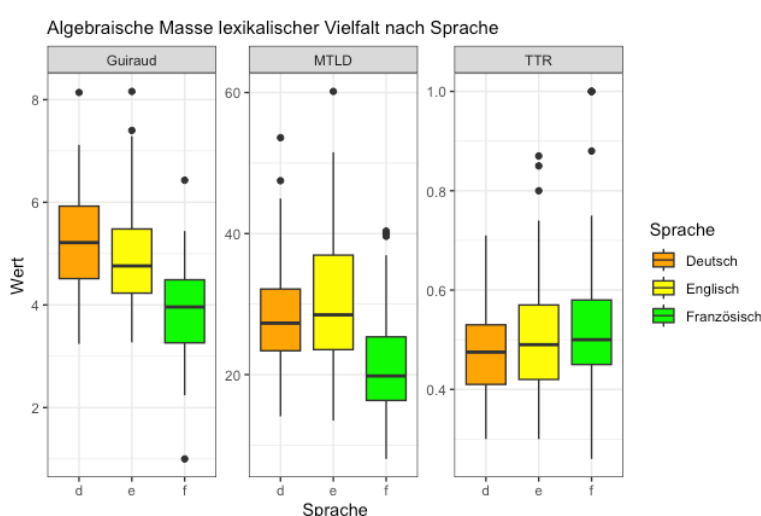
Ausgehend von den verschiedenen Massen zur Berechnung der lexikalischen Vielfalt (vgl. Kapitel 2.4.2) werden mit Excel und R Studio, nach der Zusammenführung der Daten in R Studio, die TTR, sowie der Guiraud-Index und das MTLD-Mass berechnet, wobei die verschiedenen Masse mit der lemmatisierten Originalversion (vgl. Kapitel 3.4.1.2) stark zu korrelieren scheinen, was die Korrelationskoeffizienten folgender Tabelle (vgl. Tabelle 9) ausdrücken.

Tabelle 9: Korrelationen zwischen lemmatisiertem Originaltext und bearbeiteten Versionen

Korrelation zu lemmatized_TextOrig	Deutsch			Französisch			Englisch		
	Guiraud	MTLD	TTR	Guiraud	MTLD	TTR	Guiraud	MTLD	TTR
Untersuchungsgruppe									
lemmatized_TextKorr	1.00	0.98	1.00	0.98	0.97	0.99	0.99	0.96	0.99
lemmatized_TextXXX	0.99	0.97	0.99	0.87	0.84	0.67	0.95	0.93	0.97
lemmatized_TextDelete	0.99	0.97	0.99	0.90	0.87	0.91	0.96	0.94	0.98
Vergleichsgruppe									
lemmatized_TextKorr				0.99	0.98	1.00	1.00	0.97	1.00
lemmatized_TextXXX				0.98	0.95	0.98	1.00	0.97	1.00
lemmatized_TextDelete				0.98	0.95	0.99	1.00	0.98	1.00

Die höchsten Korrelationswerte scheinen eher mit dem Guiraud-Index vorzuliegen, gefolgt von MTLD und TTR. Sowohl in der Untersuchungsgruppe als auch in der Vergleichsgruppe, liegen hohe Korrelationswerte zwischen den verschiedenen lemmatisierten Versionen mit der unkorrigierten lemmatisierten Version vor. Dies deutet darauf hin, dass die lexikalische Vielfalt in den verschiedenen Versionen in Bezug auf die lemmatisierten Originaltexte sehr ähnlich zu sein scheint. In der Vergleichsgruppe sind die Korrelationswerte jeweils etwas höher als in der Untersuchungsgruppe.

Abbildung 1: Berechnungen der lexikalischen Vielfalt der Untersuchungsgruppe mit algebraischen Massen



Die Boxplots der Abbildung 1 zeigen die Verteilung von Werten für die drei berechneten algebraischen Masse des Guiraud-Index, der MTLD und der TTR für die untersuchten Texte in

Deutsch, Englisch und Französisch der Untersuchungsgruppe. Die Boxplots sind nach Metrik aufgeteilt und zeigen für jedes algebraische Mass eine separate Boxplot-Darstellung für die drei Sprachen.

Der Wert für den Guiraud-Index und die MTLD in Englisch liegt tendenziell höher als die ermittelten Werte für Französisch. Die Werte für die deutschen Texte beim Guiraud-Index fallen tendenziell höher aus als für Englisch und Französisch. Die MTLD-Werte der deutschen Aufsätze sind zwar höher als diejenigen der französischen Aufsätze, liegen aber dennoch tendenziell unter den Werten für die englischen Textproduktionen. Bei der TTR scheinen die Werte für die deutschsprachigen Texte unter den Werten der Texte der beiden Fremdsprachen zu liegen.

Für den Guiraud-Index ist der Median für die englischsprachigen Texte (4.76) höher als in Französisch (3.96). Die deutschsprachigen Textproduktionen haben bezüglich des Guiraud-Index tendenziell höhere Werte, als die Textproduktionen in den beiden Fremdsprachen (vgl. Kapitel 8.6). Der Median für den Guiraud-Index liegt bei den deutschen Aufsätzen bei 5.22. Das erste Quartil in Englisch (4.23) liegt über dem Median in Französisch (3.26). Die Streuung nach oben ist in Englisch grösser als in Französisch. In den französischen Textproduktionen gibt es Ausreisser in beide Richtungen, während dem in den englischen Texten nur Ausreisser nach oben festzustellen sind.

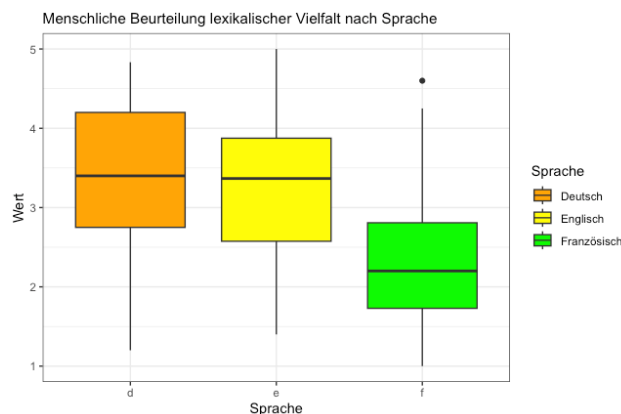
Für die MTLD ist der Median für die Texte in Englisch (28.48) höher als das obere Quartil in Französisch (25.35). Der Median für die französischen Aufsätze (19.81) liegt unterhalb des ersten Quartils (23.55) in den englischen Aufsätzen. Die Streuung nach oben ist in beiden Sprachen grösser als die Streuung nach unten. Ausreisser nach oben sind in den untersuchten Textproduktionen beider Sprachen vorhanden.

Bei der Berechnung der TTR sind die Werte in Englisch und Französisch ähnlich, wobei die Werte in Französisch minimal höher sind. Die Werte für die deutschen Texte sind tendenziell tiefer als die Werte der französischen und englischen Texte. Die Streuung insgesamt ist jedoch in Französisch grösser als in Englisch und Deutsch. Da das Mass für die TTR stark von der Textlänge abhängt (vgl. Kapitel 2.4.2.1) und die Korrelation zwischen den lemmatisierten

Originaltexten und den anderen Textversionen am geringsten ausfällt, wird in den folgenden Abschnitten mit dem Guiraud-Index und dem MTLD-Mass weitergerechnet.

Die algebraisch ermittelten Masse werden nun mit den Ergebnissen aus den Ratings der Texte durch menschliche Gutachterinnen und Gutachtern kontrastiert. Auch die Berechnungen aus den Ratings in Abbildung 2 zeigen, dass die englischen Textproduktionen in Bezug auf die lexikalische Vielfalt im Allgemeinen höher gewertet werden als die französischen Texte, dass diese zudem in einem breiteren Bereich variieren und im Median höher sind.

Abbildung 2: Rating der lexikalischen Vielfalt



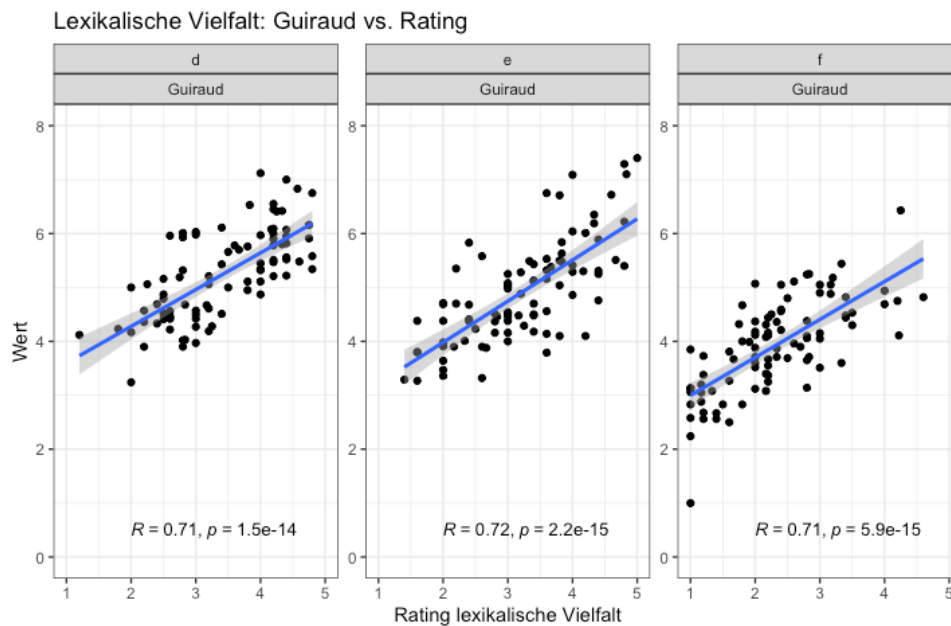
Die vorliegenden Boxplots zeigen die Verteilung der Ratings der lexikalischen Vielfalt der Texte in der Unterrichtssprache Deutsch, sowie in der ersten Fremdsprache Französisch und der zweiten Fremdsprache Englisch. Die Bewertungen wurden auf einer Skala von 1 bis 5 (1= sehr schlecht, 2= eher schlecht, 3= mittelmässig, 4= eher gut, 5= sehr gut) durchgeführt (vgl. Kapitel 3.4.2.3).

Für Englisch liegt der Median bei 3.37, was bedeutet, dass die Hälfte der Texte eine Bewertung von 3.37 oder höher erhielten. Bei den französischen Textproduktionen hingegen liegt der Median bei 2.2, was bedeutet, dass die Hälfte der Texte eine Bewertung von 2.2 oder höher erhielten. Der Median der Ratings der deutschsprachigen Texte liegt mit 3.4 nur knapp höher als der Median der französischen Textproduktionen (vgl. Kapitel 8.7).

Mit den folgenden Korrelationsdiagrammen (vgl. Abbildung 3 und 4) wird der Zusammenhang zwischen den algebraischen Massen der lexikalischen Vielfalt und den Ratings ermittelt. Die x-Achse zeigt den Mittelwert der abgegebenen Ratings pro Text auf der fünfstufigen Likert-

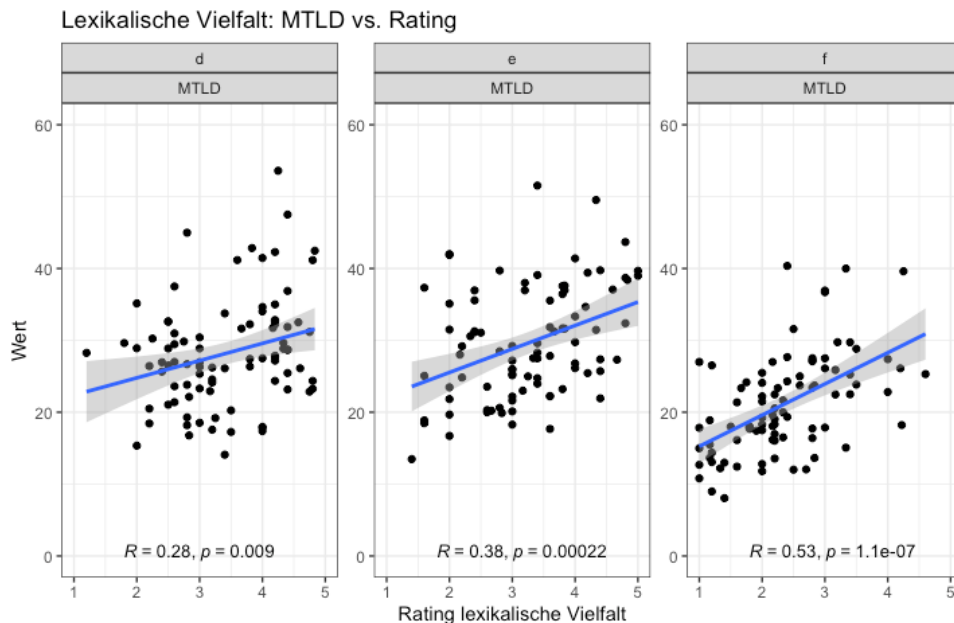
Skala. Auf der y-Achse werden die algebraisch ermittelten Werte für die Textproduktionen für den Guiraud-Index (Abbildung 3) und die MTLD (Abbildung 4) angegeben. Die Korrelationsdiagramme werden separat nach Sprache aufgeführt.

Abbildung 3: Korrelation zwischen Rating der lexikalischen Vielfalt und Guiraud-Index nach Sprache



Die drei Korrelationsdiagramme (vgl. Abbildung 3) zeigen eine positive Korrelation zwischen dem Guiraud-Index und dem Rating der lexikalischen Vielfalt für die Sprachen Deutsch, Englisch und Französisch. Der Korrelationskoeffizient (R) für jede Sprache ist ähnlich hoch und liegt zwischen 0.71 und 0.72. Dies deutet darauf hin, dass es eine starke positive Beziehung zwischen dem Guiraud-Index und dem Rating der lexikalischen Vielfalt gibt, unabhängig von der Sprache. Der p-Wert für jede Sprache ist sehr niedrig, was darauf hindeutet, dass die Korrelation statistisch signifikant ist und nicht auf Zufall zurückzuführen ist. Nachfolgend wird betrachtet, inwiefern das Rating mit der MTLD korreliert.

Abbildung 4: Korrelation zwischen Rating der lexikalischen Vielfalt und MTLD



Die Korrelationskoeffizienten und p-Werte (vgl. Abbildung 4) deuten darauf hin, dass es eine positive Korrelation zwischen dem Rating der lexikalischen Vielfalt und der MTLD in allen drei Sprachen gibt. Die Stärke der Korrelation ist jedoch unterschiedlich. Die Korrelation ist am schwächsten in Deutsch ($R=0.28$), stärker in Englisch ($R=0.38$) und am stärksten in Französisch ($R=0.53$). Die p-Werte geben an, dass die Korrelationen in allen drei Sprachen statistisch signifikant sind. Der p-Wert ist jedoch am niedrigsten in Französisch, was darauf hinweist, dass die Korrelation in Französisch am robustesten ist.

Insgesamt zeigen die Ergebnisse, dass das Rating der lexikalischen Vielfalt mit der MTLD in allen drei Sprachen korreliert. Die Ergebnisse deuten jedoch auch darauf hin, dass die Stärke der Korrelation und die Stabilität der Korrelation zwischen dem Rating der lexikalischen Vielfalt und der MTLD in verschiedenen Sprachen unterschiedlich sein können. Zudem zeigen die Abbildungen, dass bei der Gegenüberstellung zum Guiraud-Index eine Bewertung der englischen Texte in etwa der Bewertung der französischen Texte entspricht, was bei der MTLD nicht beobachtet werden kann (vgl. Kapitel 8.10).

Der Überblick über die Korrelationen zwischen dem Guiraud-Index und dem Rating, bzw. zwischen der MTLD und dem Rating in den Sprachen Deutsch, Französisch und Englisch werden nachfolgend durch den Einsatz von Regressionsmodellen untersucht. In der Tabelle 10 werden die Schätzungen für das Intercept und für den Regressionskoeffizienten für einen

Predictor sowie deren zugehörige Konfidenzintervalle und Signifikanzniveaus präsentiert. Die Anzahl der Beobachtungen für jede Analyse ist ebenfalls angegeben, sowie der Bestimmtheitsgrad (R²) und der bereinigte Bestimmtheitsgrad (bereinigtes R²) für jedes Modell.

Tabelle 10: Regressionsmodell Rating vs. algebraische Masse lexikalischer Vielfalt

Deutsch						
Predictors	Guiraud			MTLD		
	Estimates	CI	p	Estimates	CI	p
(Intercept)	2.8	2.26 – 3.33	<0.001	19.98	13.68 – 26.28	<0.001
Mittelwert LV Rating [d]	0.72	0.57 – 0.87	<0.001	2.40	0.61 – 4.18	0.009
Observations	89			89		
R ² /R ² adjusted	0.508 / 0.502			0.076 / 0.065		
Englisch						
Predictors	Guiraud			MTLD		
	Estimates	CI	p	Estimates	CI	p
(Intercept)	2.33	1.79 – 2.87	<0.001	18.88	12.74 – 25.02	<0.001
Mittelwert LV Rating [d]	0.81	0.65 – 0.97	<0.001	3.41	1.59 – 5.22	<0.001
Observations	89			89		
R ² /R ² adjusted	0.538 / 0.533			0.137 / 0.127		
Französisch						
Predictors	Guiraud			MTLD		
	Estimates	CI	p	Estimates	CI	p
(Intercept)	2.29	1.93 – 2.66	<0.001	10.91	7.19 – 14.63	<0.001
Mittelwert LV Rating [f]	0.70	0.56 – 0.85	<0.001	4.34	2.85 – 5.83	<0.001
Observations	90			87		
R ² /R ² adjusted	0.501 / 0.496			0.284 / 0.275		

Die vorliegenden linearen Regressionsanalysen (vgl. Tabelle 10) untersuchen die Beziehungen zwischen den abhängigen Variablen, welche aus den algebraisch berechneten Massen bestehen (Deutsch Guiraud, Deutsch MTLD, Französisch Guiraud, Französisch MTLD, Englisch Guiraud und Englisch MTLD) sowie den unabhängigen Variablen, welche durch die Mittelwerte der Ratings der lexikalischen Vielfalt für die deutschen, französischen und englischen Texte definiert sind. Zunächst wird das Regressionsmodell für den Zusammenhang zwischen dem Rating und dem Guiraud-Index für jede Sprache dargelegt. Anschliessend werden die Ergebnisse des Regressionsmodells zur Beziehung zwischen dem Rating und der MTLD gesondert nach Sprache aufgezeigt.

Das Regressionsmodell für Deutsch zeigt eine signifikant positive Beziehung zwischen dem Mittelwert des Ratings und dem Guiraud-Index (geschätzte Steigung von 0.72 mit einem 95%-Konfidenzintervall von 0.57 bis 0.87), wenn alle anderen Variablen konstant gehalten werden. Das Intercept von 2.8 (mit einem 95%-Konfidenzintervall von 2.26 bis 3.33) kann als der erwartete Wert des Guiraud-Index interpretiert werden, wenn das Rating gleich Null ist.

Das Regressionsmodell für Englisch zeigt auch eine signifikant positive Beziehung zwischen dem Mittelwert des Ratings und dem Guiraud-Index (geschätzte Steigung von 0.81 mit einem 95%-Konfidenzintervall von 0.65 bis 0.97) und ein Intercept von 2.33 (mit einem 95%-Konfidenzintervall von 1.79 bis 2.87).

Ebenfalls das Regressionsmodell für Französisch zeigt eine signifikant positive Beziehung zwischen dem Mittelwert des Ratings und dem Guiraud-Index (geschätzte Steigung von 0.7 mit einem 95%-Konfidenzintervall von 0.56 bis 0.85) und ein Intercept von 2.29 (mit einem 95%-Konfidenzintervall von 1.93 bis 2.66).

Insgesamt zeigen alle Regressionsmodelle eine positive Beziehung zwischen dem Rating und dem Guiraud-Index, was bedeutet, dass Texte mit höheren Ratings tendenziell höhere Guiraud-Indizes aufweisen. Der R²-Wert gibt an, dass etwa die Hälfte der Varianz im Guiraud-Index durch den Mittelwert des Ratings erklärt werden kann (0.508 für Deutsch, 0.538 für Englisch und 0.501 für Französisch).

Werden die Werte des MTLD betrachtet, zeigt das Regressionsmodell für Deutsch eine signifikant positive Beziehung zwischen dem Mittelwert des Ratings und dem MTLD (geschätzte Steigung von 2.4 mit einem 95%-Konfidenzintervall von 0.61 bis 4.18), wenn alle anderen Variablen konstant gehalten werden. Das Intercept von 19.98 (mit einem 95%-Konfidenzintervall von 13.68 bis 26.28) kann als der erwartete Wert des MTLD interpretiert werden, wenn das Rating gleich Null ist. Der R²-Wert gibt an, dass nur ein kleiner Teil der Varianz des MTLD durch den Mittelwert des Ratings erklärt werden kann (0.0076).

Das Regressionsmodell für Englisch zeigt eine signifikant positive Beziehung zwischen dem Mittelwert des Ratings der lexikalischen Vielfalt und der MTLD (geschätzte Steigung von 3.41 mit einem 95%-Konfidenzintervall von 1.59 bis 5.22) und ein Intercept von 18.88 (mit einem 95%-Konfidenzintervall von 12.74 bis 25.02). Der R²-Wert gibt an, dass etwa 13.7% der Varianz der MTLD durch den Mittelwert des Ratings erklärt werden können.

Auch für die französischen Textproduktionen zeigt die Analyse eine signifikant positive Beziehung zwischen dem Mittelwert des Ratings der lexikalischen Vielfalt und der MTLD (geschätzte Steigung von 4.34 bei einem 95%-Konfidenzintervall von 2.85-5.83), sowie ein Intercept von 10.91 (mit einem 95%-Konfidenzintervall von 7.19 – 14.63). Das Regressionsmodell erklärt 28.4% der Gesamtvariation der MTLD in der Sprache Französisch, welche durch Veränderungen im Rating verursacht werden.

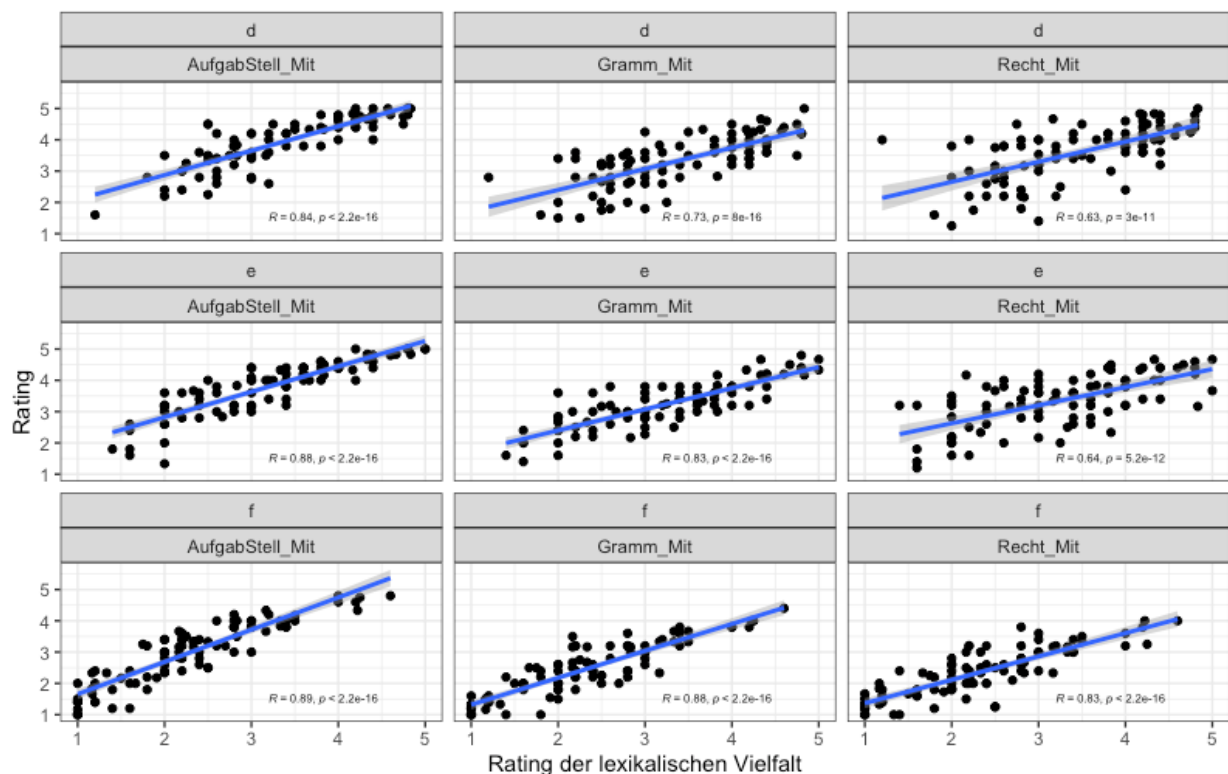
Zusammenfassend zeigen diese Regressionsanalysen, dass das Rating in Bezug auf die lexikalische Vielfalt der Textproduktionen eine positive Beziehung zur abhängigen Variable aufweist, wobei das Rating für die Texte in Englisch höhere Schätzungen als das Rating für die französischen Texte zeigt.

Insgesamt bestätigen die Ergebnisse der Regressionsanalysen die in Abbildung 3 und 4 beobachteten Korrelationen, indem höhere Ratings mit höheren algebraisch ermittelten Werten für den Guiraud-Index und die MTLD in beiden Sprachen korrelieren. Die Modelle zeigen aber, dass die MTLD als Indikator für die lexikalische Vielfalt eines Textes im Vergleich zum Guiraud-Index eine deutlich geringere Vorhersagekraft zu haben scheint. Die R²-Werte für die Regressionsmodelle mit der MTLD liegen zwischen 0.0076 und 0.284, während die R²-Werte für die Regressionsmodelle mit dem Guiraud-Index zwischen 0.501 und 0.538 liegen. Das bedeutet, dass der Guiraud-Index tendenziell eine bessere Vorhersage für die lexikalische Vielfalt eines Textes liefert als die MTLD-Werte.

In Bezug auf die Sprachen kann beobachtet werden, dass der Guiraud-Index für Englisch die höchste Vorhersagekraft hat, gefolgt von Deutsch und Französisch. Bei den MTLD-Modellen zeigt sich jedoch ein anderes Bild. Hier hat Französisch die höchste Vorhersagekraft, gefolgt von Deutsch und Englisch. Es gilt aber zu beachten, dass der R²-Wert bei allen Modellen relativ niedrig ist, was bedeutet, dass die Vorhersagekraft insgesamt begrenzt ist und es weitere Faktoren gibt, die die lexikalische Vielfalt des Textes beeinflussen.

Die menschlichen Gutachterinnen und Gutachter haben die Textproduktionen in den verschiedenen Sprachen ebenfalls in Bezug auf die Aufgabenstellung, die Grammatik und die Rechtschreibung beurteilt. In der Abbildung 5 wird der Zusammenhang zwischen dem Rating der lexikalischen Vielfalt und den Ratings zur Aufgabenstellung, zur Grammatik und zur Rechtschreibung, gesondert nach Sprache aufgezeigt.

Abbildung 5: Korrelation Rating der lexikalischen Vielfalt mit Aufgabenstellung, Grammatik und Rechtschreibung



Der Korrelationskoeffizient gibt an, wie stark die Beziehung zwischen dem Rating der lexikalischen Vielfalt der Texte und den entsprechenden Variablen ist. Der positive Korrelationskoeffizient der verschiedenen Streudiagramme wird mit der Regressionsgeraden dargestellt und bedeutet, dass die Variablen jeweils mit dem Rating der lexikalischen Vielfalt positiv korreliert sind.

Die Ergebnisse zeigen starke positive Korrelationen zwischen dem Rating der lexikalischen Vielfalt und dem Rating der untersuchten Kriterien (Aufgabenstellung, Grammatik, Rechtschreibung) in allen drei Sprachen. Die Korrelationskoeffizienten zwischen dem Rating der lexikalischen Vielfalt und dem Rating der Grammatik sind in allen drei Sprachen hoch, wobei die Korrelation in Französisch ($R=0.88$) am höchsten ist, gefolgt von Englisch ($R=0.83$) und Deutsch ($R=0.73$). Die Korrelationskoeffizienten zwischen dem Rating der lexikalischen Vielfalt und dem Rating der Rechtschreibung sind ebenfalls in allen Sprachen positiv, wobei die höchste Korrelation, im Vergleich zu Deutsch ($R=0.63$) und Englisch ($R=0.64$), ebenfalls in Französisch ($R=0.83$) auftritt. Werden die Werte der Ratings der lexikalischen Vielfalt den Ratings bezüglich Aufgabenstellung gegenübergestellt, sind die Korrelationskoeffizienten in

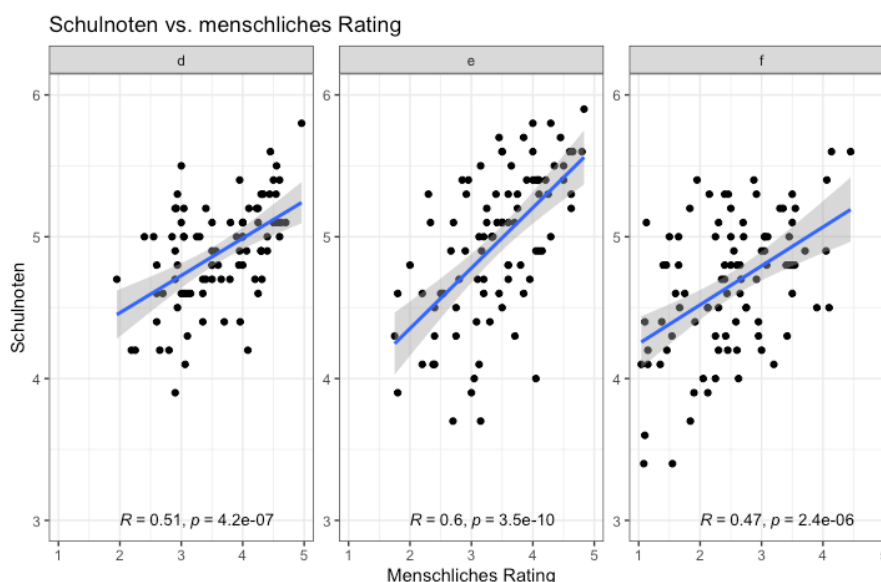
allen drei untersuchten Sprachen hoch, wobei wiederholt in Französisch der höchste Korrelationswert ($R=0.89$) vorliegt (Deutsch $R=0.84$, Englisch $R=0.88$).

Im Allgemeinen sind die Korrelationen der Ratings der lexikalischen Vielfalt in den Französischtexten und den Ratings der Aufgabenstellung, der Grammatik und der Rechtschreibung am höchsten, gefolgt von Englisch und Deutsch. Der Zusammenhang zwischen dem Rating der lexikalischen Vielfalt und dem Rating der Grammatik ist in allen drei Sprachen ähnlich stark. Ebenso ist der Zusammenhang zwischen dem Rating der lexikalischen Vielfalt und der Rechtschreibung der Aufsätze in Deutsch und Englisch ähnlich stark, während dem die Korrelation in den Französischtexten am stärksten ausfällt.

Die p-Werte zeigen, dass die Korrelationen zwischen den Ratings der lexikalischen Vielfalt und den Ratings der untersuchten Kriterien zu den Sprachfertigkeiten signifikant sind und nicht zufällig auftreten.

Als zusätzliche Variable wurden bei den Lehrpersonen die Notendurchschnitte im Bereich Schreiben in den jeweiligen Sprachen erhoben (vgl. Kapitel 8.3). In der folgenden Abbildung werden die Mittelwerte der menschlichen Ratings, berechnet aus dem Mittelwert der vier Parameter (lexikalische Vielfalt, Aufgabenstellung, Grammatik und Rechtschreibung) den Notendurchschnitten in der jeweiligen Sprache im Bereich Schreiben gegenübergestellt.

Abbildung 6: Korrelation Rating der lexikalischen Vielfalt und Schulnoten



Die Abbildung 6 deutet mit der eingezeichneten Regressionsgeraden darauf hin, dass in allen drei untersuchten Sprachen eine positive Korrelation zwischen den Notendurchschnitten im Bereich Schreiben und dem Mittelwert der Ratings der untersuchten Texte besteht. In Bezug auf die Streuung ist ersichtlich, dass in den Sprachen Deutsch und Englisch die Werte für das menschliche Rating und den Notendurchschnitt höher sind als in der Sprache Französisch. Das Streudiagramm zeigt zudem auf, dass in Französisch die Werte im Rating und die Schulnoten tiefer angesiedelt sind als in Englisch und Deutsch. Mit der Tabelle 11 werden die linearen Regressionsanalysen durchgeführt und anschliessend beschrieben.

Tabelle 11: Regressionsanalyse Rating vs. Schulnoten nach Sprachen

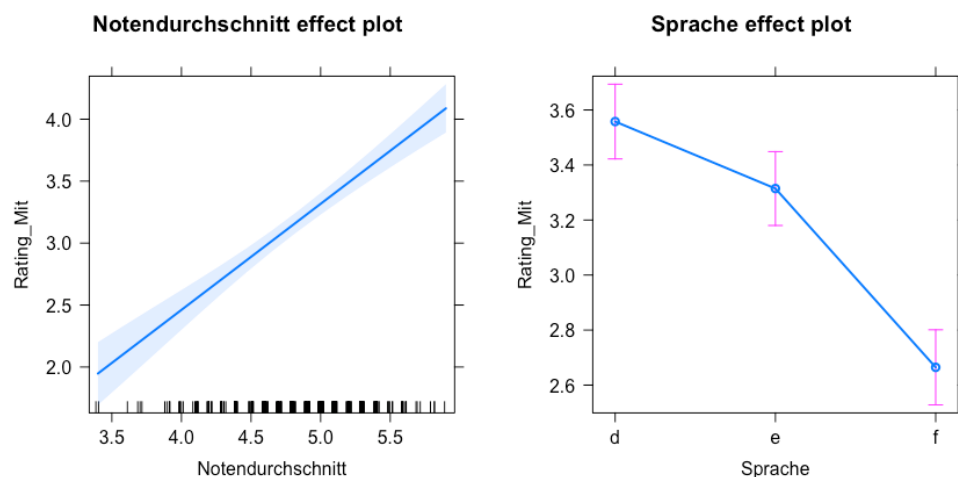
Predictors	Rating Mittelwert		
	Estimates	CI	p
(Intercept)	-0.58	-1.41 – 0.25	0.171
Notendurchschnitt	0.86	0.69 – 1.02	< 0.001
Französisch	-0.89	-1.09 – -0.70	0.012
Englisch	-0.24	-0.43 – -0.05	< 0.001
Observations	271		
R ² /R ² adjusted	0.480 / 0.474		

Die Tabelle 11 zeigt Berechnungen der linearen Regression für die Untersuchungsgruppe. Im Regressionsmodell wird eine lineare Regression durchgeführt, in welcher der Mittelwert der Ratings, bestehend aus den Ratings der lexikalischen Vielfalt, der Ratings in Bezug auf die Aufgabenstellung, sowie bezüglich Grammatik und Rechtschreibung, als abhängige Variable und der Notendurchschnitt sowie die Sprache als unabhängige Variablen betrachtet werden.

Die Korrelation zwischen dem Rating und dem Notendurchschnitt, unabhängig von der Sprache, wird im Notendurchschnitt effect plot durch die Steigung dargestellt (vgl. Abbildung 7). Das schmale Konfidenzband kann ein Hinweis darauf sein, dass die Standardabweichung gering ist. Der Sprache effect plot (vgl. Abbildung 7) zeigt, wie sich die Ratings der Texte in Abhängigkeit von der Sprache der Texte unterscheiden. Das Modell zeigt, dass es einen signifikanten Einfluss der Sprache auf die Ratings gibt, da die p-Werte für Englisch und Französisch kleiner als 0,05 sind (vgl. Tabelle 11). Der Effekt der Sprache wird in Bezug auf die Referenzkategorie Deutsch interpretiert. Der Schätzwert (Estimate) von -0.24 für Englisch besagt, dass das durchschnittliche Rating der englischen Texte um 0.24 Punkte niedriger ist

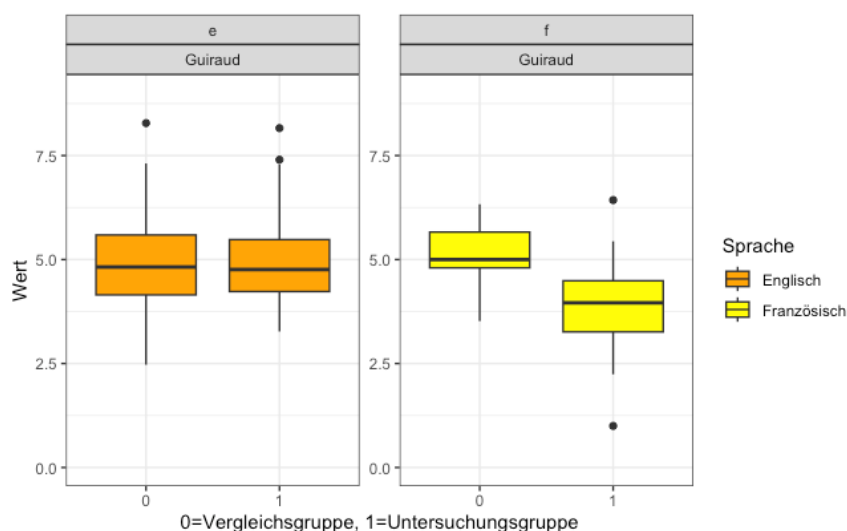
als das der deutschen Texte, wenn der Notendurchschnitt und die Interventionsgruppe konstant gehalten werden. Im Vergleich dazu liegt der Schätzwert für Französisch bei -0.89, was darauf hindeutet, dass das durchschnittliche Rating der französischen Texte um 0.89 Punkte niedriger ist als das der deutschen Texte, wenn der Notendurchschnitt und die Interventionsgruppe konstant gehalten werden. Der Sprache effect plot (vgl. Abbildung 7) verdeutlicht mit dem eingezeichneten Konfidenzintervall, welches durch den Bereich um den beobachtbaren Mittelwert der Ratings definiert ist und innerhalb von welchem der wahre Mittelwert mit einer Wahrscheinlichkeit von 95% liegt, dass der Unterschied zwischen den Textproduktionen der zweiten Fremdsprache Englisch und Deutsch kleiner ist als der Unterschied zwischen den Texten der beiden Fremdsprachen, sowie den deutschen und französischen Texten.

Abbildung 7: Effect plots



Aufgrund typologischer Unterschiede zwischen den Sprachen Englisch und Französisch sind, wie im Kapitel 2.4.2 erläutert, direkte Vergleiche zwischen Sprachen nicht möglich. Um den Unterschied zwischen der Untersuchungsgruppe und einer Vergleichsgruppe hinsichtlich ihrer lexikalischen Vielfalt zu verdeutlichen, werden in Abbildung 8 der Guiraud-Index und die MTLD-Werte beider Gruppen gegenübergestellt. Das Ziel besteht darin, die Grösse des Unterschieds zwischen der Untersuchungsgruppe, welche Französisch und Englisch als erste und zweite Fremdsprache in der Schule erlernen, und den Vergleichsgruppen, für welche Französisch oder Englisch die Unterrichtssprache ist, aufzuzeigen.

Abbildung 8: Vergleich Guiraud-Index Untersuchungsgruppe vs. Vergleichsgruppe



Der Boxplot in Abbildung 8 zeigt die Verteilung des Guiraud-Index für die Sprachen Englisch und Französisch in der Vergleichs- und der Untersuchungsgruppe. Für Englisch zeigt der Boxplot, dass die Untersuchungsgruppe im Median einen etwas niedrigeren Guiraud-Index (4.76) als die Vergleichsgruppe (4.82) aufweist (vgl. Kapitel 8.8). Der Interquartilsabstand, der durch die Differenz zwischen dem ersten und dritten Quartil bestimmt wird, ist in beiden Gruppen sehr ähnlich, was darauf hindeutet, dass die Verteilung in beiden Gruppen recht ähnlich ist.

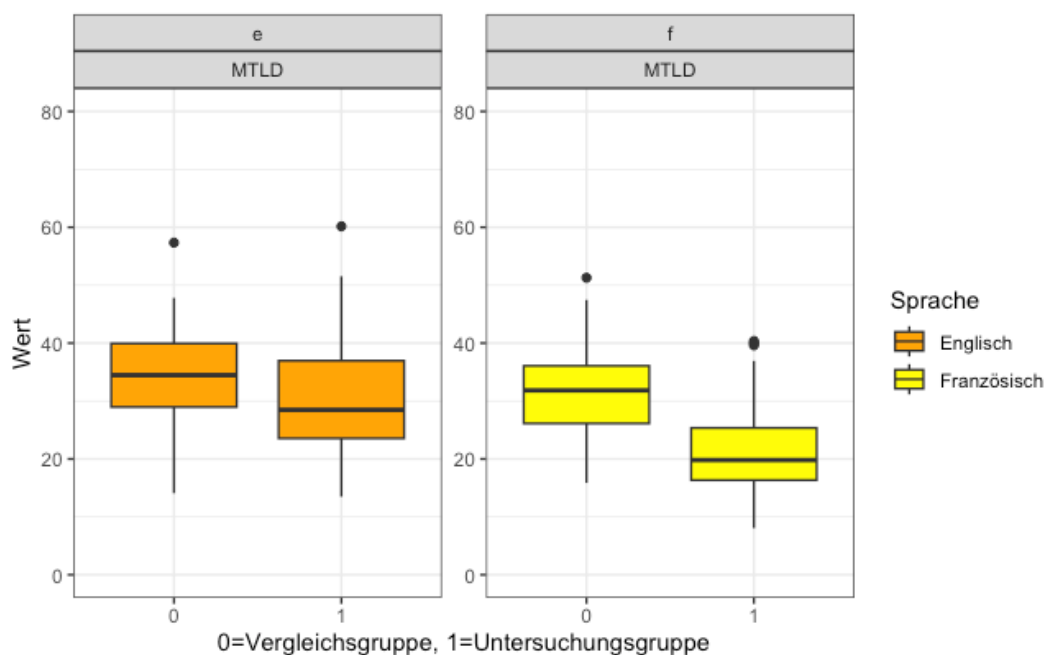
Für Französisch zeigt der Boxplot, dass die Untersuchungsgruppe im Median einen deutlich niedrigeren Guiraud-Index (3.96) als die Vergleichsgruppe (5) aufweist. Der Interquartilsabstand ist in der Untersuchungsgruppe etwas grösser als in der Vergleichsgruppe, was darauf hindeutet, dass die Daten in der Vergleichsgruppe stärker um den Median gruppiert sind als in der Untersuchungsgruppe. Insgesamt zeigt der Boxplot somit, dass die lexikalische Vielfalt in der Untersuchungsgruppe im Median niedriger ist als in der Vergleichsgruppe.

Die Untersuchungsgruppen haben insgesamt niedrigere Guiraud-Index-Werte als die Vergleichsgruppen. Die Medianwerte liegen bei 4,76 im Englischen und 3,96 im Französischen. Die Tendenz ist dahingehend, dass die Textproduktionen der Schülerinnen und Schüler der Untersuchungsgruppe in beiden Fremdsprachen einen geringeren Wert für die lexikalische

Vielfalt aufweisen als die Texte der Lernenden der Vergleichsgruppe, wobei der Unterschied bei den französischen Aufsätzen grösser ist als bei den englischen Textproduktionen.

Beim Vergleich der ermittelten Werte des Guiraud-Index mit den MTLD-Werten sind dieselben Tendenzen auszumachen (vgl. Abbildung 9). Die Boxplots in Abbildung 9 zeigen die Verteilung der MTLD-Werte für die Untersuchungs- und Vergleichsgruppe in der jeweiligen Sprache Französisch und Englisch.

Abbildung 9: Vergleich MTLD Untersuchungsgruppe vs. Vergleichsgruppe



In der englischen Sprache zeigt der Boxplot eine höhere Streuung in der Vergleichsgruppe im Vergleich zur Untersuchungsgruppe. Der Median in der Vergleichsgruppe ist höher als in der Untersuchungsgruppe, was darauf hinweist, dass die MTLD-Werte für die lexikalische Vielfalt der englischen Sprache in der Vergleichsgruppe höher sind. Der Interquartilsabstand, der die mittleren 50% der Daten um den Median herum darstellt, ist auch in der Vergleichsgruppe grösser, was auf eine höhere Varianz der MTLD in dieser Gruppe hinweist (vgl. Kapitel 8.9).

In der französischen Sprache zeigt der Boxplot einen noch grösseren Unterschied zwischen den beiden Gruppen. Die Unterschiede zwischen den Gruppen sind in Französisch deutlicher

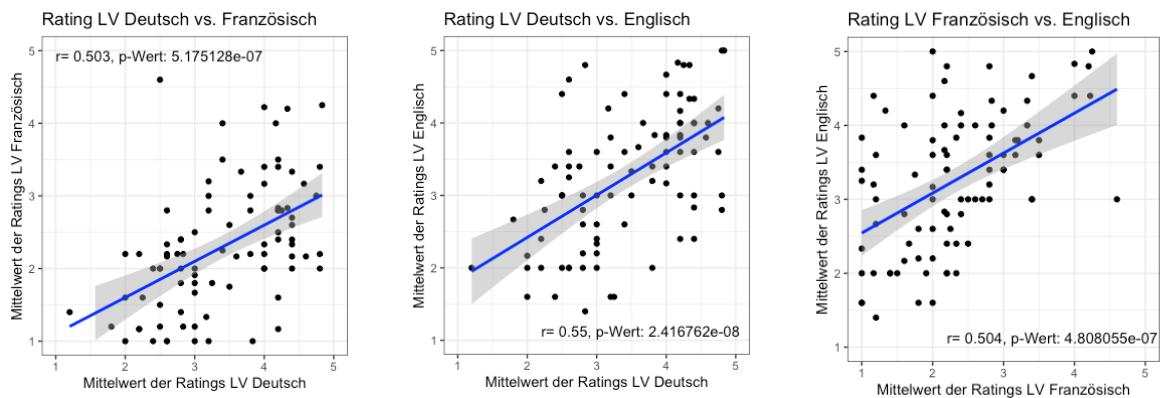
als in Englisch mit einem deutlich niedrigeren Median und Interquartilsabstand in der Untersuchungsgruppe im Vergleich zur Vergleichsgruppe.

Insgesamt deuten die Kennwerte darauf hin, dass die lexikalische Vielfalt in der Untersuchungsgruppe in beiden Sprachen niedriger ist als in der Vergleichsgruppe, wobei die ermittelten Werte für die MTLD bei den französischen Textproduktionen der Untersuchungsgruppe stärker von der Vergleichsgruppe abweichen als bei den englischen Aufsätzen.

Im Anschluss wird untersucht, inwiefern ein Zusammenhang zwischen den mittels Rating beurteilten Texten in den verschiedenen Sprachen besteht. Dazu werden die Daten der Untersuchungsgruppe verwendet, um die Korrelation der Daten zu berechnen (vgl. Tabelle 11), wobei diese paarweise verglichen werden. Das bedeutet, dass nur die Datenpaare berücksichtigt werden, die vollständig sind und keine fehlenden Werte haben. Folglich beziehen sich die Berechnungen auf diejenigen Lernenden, welche in allen drei Sprachen einen Aufsatz geschrieben haben.

Die Abbildung 10 zeigt die Korrelationskoeffizienten, welche aus den berechneten Mittelwerten der im Rating beurteilten lexikalischen Vielfalt der Texte in Deutsch, Französisch und Englisch, wie sie von den Gutachterinnen und Gutachtern in der Untersuchungsgruppe wahrgenommen werden, und der Gegenüberstellung der Sprachen berechnet werden. Die Korrelationskoeffizienten zwischen Deutsch und Französisch, Deutsch und Englisch sowie Französisch und Englisch sind positiv, was vermuten lässt, dass es eine positive lineare Beziehung zwischen der lexikalischen Vielfalt in diesen Sprachen gibt.

Abbildung 10: Korrelationsdiagramme Zusammenhang lexikalische Vielfalt zwischen den Sprachen



Die höchste Korrelation besteht zwischen Deutsch und Englisch mit einem Wert von $r=0.55$, gefolgt von Französisch und Englisch, bzw. Deutsch und Französisch, die einen ähnlichen Korrelationskoeffizienten ($r=0.503$, bzw. $r=0.504$) aufweisen. Die Korrelationskoeffizienten deuten darauf hin, dass ein Zusammenhang zwischen der lexikalischen Vielfalt der Texte der Unterrichtssprache und den beiden Fremdsprachen besteht und die Textproduktionen in Deutsch und Englisch die stärkste Beziehung zueinander haben. Das bedeutet, dass Personen, die in Deutsch über eine grössere lexikalische Vielfalt im Bereich Schreiben verfügen, tendenziell auch mehr Wörter in Englisch und Französisch kennen.

4 Diskussion

Die vorliegende Diskussion soll zeigen, inwiefern sich die lexikalische Vielfalt in Textproduktionen von Lernenden am Ende der obligatorischen Schulzeit in der ersten unterrichteten Fremdsprache Französisch zu der zweiten unterrichteten Fremdsprache Englisch unterscheidet. Zunächst werden die Ergebnisse der Erhebung mit Bezugnahme zu den Hypothesen, welche der Arbeit zugrunde liegen (vgl. Kapitel 3.2), dargestellt. In diesem Abschnitt soll nicht erneut auf alle Details der Ergebnispräsentation eingegangen werden. Die spezifischen Ergebnisse werden im empirischen Teil (vgl. Kapitel 3) detailliert erläutert und durch Abbildungen und Tabellen veranschaulicht. Stattdessen sollen hier die wichtigsten Erkenntnisse der Untersuchung zusammenfassend betrachtet und interpretiert werden, um einen Überblick zu geben.

4.1 Hypothese 1

Die Auswertung der vorliegenden Untersuchung gibt Hinweise darauf, dass die erste Hypothese, wonach die Textproduktionen von Lernenden am Ende der obligatorischen Schulzeit in der zweiten Fremdsprache Englisch eine höhere algebraisch berechnete lexikalische Vielfalt aufweisen als ihre Textproduktionen in der ersten Fremdsprache Französisch, tendenziell unterstützt wird. In der Untersuchung werden drei Masse für die Berechnung der lexikalischen Vielfalt, nämlich die TTR, der Guiraud-Index und die MTLD berücksichtigt.

Bei der TTR scheinen die Werte für die deutschen Texte unter den Werten der Texte in den beiden Fremdsprachen zu liegen. Obwohl die TTR zwar, wie im Kapitel 2.4.2.1 erwähnt, intuitiv hilfreich sein kann, ist sie anfällig für die Textgrösse und eignet sich nur begrenzt für den Vergleich von Texten unterschiedlicher Länge. Wenn ein Text länger wird, sinkt die TTR, da es weniger neue Types gibt und bereits vorhandene Types häufiger wiederholt werden (McCarthy & Jarvis, 2007, S. 60). Die TTR benachteiligt daher längere Texte (Mellor, 2011, S. 3), weshalb dieses Mass der lexikalischen Vielfalt in dieser Untersuchung wenig aussagekräftig ist und nicht weiter berücksichtigt wird.

Eine Tendenz, dass die lexikalische Vielfalt in den englischen Textproduktionen grösser ist als in den französischen Aufsätzen zeigen die Analysen der Boxplots (vgl. Abbildung 1), bei welchen die Werte für die MTLD in Englisch tendenziell höher liegen als die Werte für Französisch. Auffallend ist, dass die MTLD-Werte für die deutschen Textproduktionen zwar höher sind als für Französisch, aber dennoch tendenziell unter den Werten für die englischen Textproduktionen liegen. Obwohl die MTLD über weite Bereiche der Stichprobengrösse relativ konstant zu sein scheint (vgl. Covington & McFall, 2010; Malvern et al., 2004; McCarthy & Jarvis, 2010), zeigt Koizumi (2012, S. 67) auf, dass die MTLD empfindlich für Stichprobeneffekte bei kurzen Texten mit 50 bis 200 Tokens ist. In der vorliegenden Untersuchung schwankt die Anzahl der Tokens der Textproduktionen in den deutschen Texten zwischen 34 und 362 Tokens, in den französischen Textproduktionen zwischen 8 und 223 Tokens und in den englischen Aufsätzen zwischen 15 und 380 Tokens. Dabei können die Empfehlungen von Koizumi (2012) nicht berücksichtigt werden, die besagen, dass die Texte zur Berechnung der MTLD mindestens 100 Tokens beinhalten und die maximale Differenz von 50 Tokens nicht überschreiten sollen. Auch Bonvin et al. (2018, S. 137) deuten zwar darauf hin, dass die MTLD sich als Mass für die Berechnung der lexikalischen Vielfalt von nahezu allen Texten anbietet, obwohl diese bei kurzen Texten nicht empfohlen wird.

Die Berechnungen des Guiraud-Index zur Bestimmung der lexikalischen Vielfalt der gesammelten Texte zeigen die Tendenz, dass die englischen Textproduktionen eine höhere lexikalische Vielfalt aufweisen als die Textproduktionen in der ersten Fremdsprache Französisch, am deutlichsten (vgl. Abbildung 1). Wie eingangs erwähnt (Kapitel 2.4.2.3), scheint der Guiraud-Index für Textumfänge von mehreren hundert Wörtern stabil zu sein (Daller, 2010, S. 12) und zudem ergaben verschiedene Messungen, dass der Guiraud-Index der wichtigste Prädiktor für die Beurteilung des lexikalischen Reichtums ist (Vanhove et al., 2019, S. 516; Kuhn & Johnson, 2013), was in den folgenden Abschnitten näher betrachtet wird.

4.2 Hypothese 2

Um die Hypothese zu überprüfen, ob die lexikalische Vielfalt in den englischen Textproduktionen von menschlichen Ratern als höher eingestuft wird als in den französischen Textproduktionen, können die Medianwerte aus den Bewertungen mit der fünfstufigen Likert-Skala herangezogen werden. Der Medianwert der englischen Textproduktionen beträgt 3.37,

während der Medianwert der französischen Textproduktionen bei 2.2 liegt (vgl. Abbildung 2). Es zeigt sich, dass die menschlichen Raterinnen und Rater die englischen Texte als lexikalisch vielfältiger bewerten als die französischen Texte. Zudem sind auch die Quartilswerte (Q1-Q3) von Bedeutung, um weitere Einblicke in die Bewertung der lexikalischen Vielfalt zu erhalten. Die englischen Textproduktionen weisen hier einen höheren Bereich auf (Q1=2.6, Q3=3.9) im Vergleich zu den französischen Textproduktionen (Q1=1.7, Q3=2.8), was ebenfalls auf eine höhere lexikalische Vielfalt in den englischen Texten hinweist. Bei der Betrachtung der Boxplots (vgl. Abbildung 2) fällt auf, dass die Bewertungen der Texte in der Unterrichtssprache Deutsch mit einem Median von 3.4 (Q1=2.8, Q3=4.2) nur knapp oberhalb der Bewertungen der Texte in der zweiten Fremdsprache Englisch liegen, während die Beurteilungen der Texte in der ersten Fremdsprache Französisch deutlich niedriger ausfallen.

Dies könnte auf verschiedene Faktoren zurückzuführen sein. Zum einen können die zukünftigen Lehrpersonen des Ratings möglicherweise höhere Erwartungen an die sprachliche Kompetenz ihrer Schülerinnen und Schüler in der Unterrichtssprache haben, da diese Sprache für den Unterricht und für das Verständnis von Lerninhalten unerlässlich ist. Darüber hinaus kann es auch sein, dass die angehenden Lehrpersonen, welche Deutsch als Muttersprache angegeben haben (vgl. Tabelle 8), aufgrund ihrer eigenen sprachlichen Kompetenz in der Unterrichtssprache eine präzisere Beurteilung der lexikalischen Vielfalt vornehmen können. Auf der anderen Seite beurteilen die Lehrpersonen in Ausbildung die lexikalische Vielfalt von Texten in einer Fremdsprache möglicherweise grosszügiger, da sie sich bewusst sind, dass die Lernenden in den Fremdsprachen nicht über die gleiche lexikalische Vielfalt wie in der Unterrichtssprache verfügen. Darüber hinaus besteht die Möglichkeit, dass die Gutachterinnen und Gutachter in der Fremdsprache möglicherweise weniger genau oder weniger schnell Fehler erkennen, da sie selbst nicht in der Fremdsprache aufgewachsen sind oder diese nicht auf dem Niveau eines Muttersprachlers beherrschen.

Um zu untersuchen, inwiefern die Ratings mit den algebraisch berechneten Indizes der lexikalischen Vielfalt in den verschiedenen Sprachen korrelieren, wird in dieser Arbeit eine Korrelationsanalyse durchgeführt. Im Folgenden wird die dritte Hypothese vorgestellt zur Korrelation zwischen den algebraischen Massen der lexikalischen Vielfalt und den Ergebnissen aus den Ratings in den verschiedenen Sprachen dargelegt.

4.3 Hypothese 3

Gemäss den Untersuchungen korrelieren die algebraischen Masse der lexikalischen Vielfalt mit den Ergebnissen aus den Ratings nicht nur in der ersten und zweiten Fremdsprache, sondern auch in der Unterrichtssprache, insbesondere in Bezug auf den Guiraud-Index. Die Korrelationsdiagramme in Abbildung 3 zeigen eine positive Korrelation zwischen dem Guiraud-Index und dem Rating der lexikalischen Vielfalt sowohl in der Unterrichtssprache Deutsch als auch in der ersten Fremdsprache Französisch und der zweiten Fremdsprache Englisch. Diese hohen Korrelationskoeffizienten für jede Sprache, welche dicht beieinander liegen (zwischen 0.71 und 0.72), könnte als Evidenz dafür verwendet werden, dass eine höhere lexikalische Vielfalt mit einem höheren Guiraud-Index einhergeht, und dass diese Beziehung für alle drei getesteten Sprachen gilt. Die niedrigen p-Werte deuten zudem darauf hin, dass diese Korrelationen statistisch signifikant sind und nicht auf Zufall zurückzuführen sind. Der Guiraud-Index und das Rating der lexikalischen Vielfalt, bzw. das Rating in Bezug auf die lexikalische Vielfalt, die Aufgabenstellung, die Grammatik und die Rechtschreibung, könnten daher als nützliche Indikatoren für die Sprachkompetenz im Bereich Schreiben verwendet werden.

Anders sieht es mit der Korrelation zwischen dem MTLD-Mass und dem Rating aus (vgl. Abbildung 4). Zwar deuten die Korrelationskoeffizienten und die p—Werte darauf hin, dass es eine positive Korrelation zwischen dem Rating der lexikalischen Vielfalt und der MTLD in allen drei Sprachen gibt. Jedoch ist die Korrelation weniger ausgeprägt als beim Guiraud-Index und variiert stärker zwischen der Unterrichtssprache Deutsch ($r=0.28$), der ersten Fremdsprache Französisch ($r=0.53$) und der zweiten Fremdsprache Englisch ($r=0.38$). Mögliche Gründe dafür werden in der Auswertung der ersten Hypothese bereits diskutiert (vgl. Kapitel 4.1).

4.4 Hypothese 4

Eine weitere Hypothese besagt, dass die Wahrnehmung der lexikalischen Vielfalt in der ersten und zweiten Fremdsprache mit der Wahrnehmung der Erreichung der Aufgabenstellung, der Grammatik und der Rechtschreibung der Textproduktionen korreliert. Die Ergebnisse der Untersuchung bestätigen diese Hypothese, da starke positive Korrelationen zwischen dem Rating der lexikalischen Vielfalt und den Ratings zur Aufgabenstellung, zur Grammatik und zur Rechtschreibung in allen drei Sprachen beobachtet werden können (vgl. Abbildung 5). Die

höchsten Korrelationswerte zwischen dem Rating der lexikalischen Vielfalt und den Ratings der untersuchten Kriterien liegen in den französischen Texten vor, gefolgt von den englischen und deutschen Textproduktionen. Die Korrelationen zwischen dem Rating der lexikalischen Vielfalt und der Grammatik sind in allen drei Sprachen ähnlich stark (zwischen 0.73 und 0.88), während die Korrelationen zwischen dem Rating der lexikalischen Vielfalt und der Rechtschreibung in Französisch ($r=0.83$) im Vergleich zu den englischen Texten ($r=0.64$) und den deutschen Textproduktionen ($r=0.63$) am stärksten ausfallen. Die Ergebnisse zeigen (vgl. Abbildung 5) im Allgemeinen eine geringere Varianz und eine stärkere Konzentration der Bewertungen im unteren Bereich für die französischen Texte im Vergleich zu den deutschen und englischen Texten, bei denen eine höhere Varianz und eine stärkere Konzentration der Bewertungen im oberen Bereich zu beobachten sind. Die p-Werte deuten darauf hin, dass diese Korrelationen signifikant sind und nicht zufällig auftreten.

4.5 Hypothese 5

Die Hypothese legt dar, dass die Wahrnehmung der lexikalischen Vielfalt, der Erreichung der Aufgabenstellung, der Grammatik und der Rechtschreibung mit dem Notendurchschnitt der Lernenden am Ende der obligatorischen Schulzeit im Bereich Schreiben in der ersten und zweiten Fremdsprache korreliert.

Gemäss den Ergebnissen aus der Untersuchung ergibt sich, dass es in allen drei Sprachen (Deutsch, Englisch und Französisch) eine positive Korrelation zwischen den Notendurchschnitten im Bereich Schreiben und dem Mittelwert der Ratings der untersuchten Texte gibt. Die linearen Regressionsanalysen bestätigen diese Korrelation und verdeutlichen, dass der Notendurchschnitt und die Sprache einen signifikanten Einfluss auf die Ratings haben. Es ist interessant zu sehen, dass die Unterschiede zwischen den Textproduktionen der zweiten Fremdsprache Englisch und Deutsch geringer sind als die Unterschiede zwischen den Texten der beiden Fremdsprachen sowie den deutschen und französischen Texten (vgl. Abbildung 7).

Die Gründe können vielfältig sein: Die typologische Spezifität der Sprache (Treffers-Daller et al., 2016, S. 11), Ergebnisse aus der Motivationsforschung (z.B. Stöckli, 2004; Heinzmann, 2013; Peyer et al., 2016a) und bildungspolitische Aspekte rund um die Debatte bezüglich der

Reihenfolge der Einführung der Fremdsprachen (z.B. Stöckli, 2004; HarmoS-Konkordat, 2008; Peyer et al., 2016b), die Bedeutung der Mehrsprachigkeitsdidaktik, Fragen bezüglich des Altersfaktors in Bezug auf den Lernerfolg (z.B. Lambelet & Berthele, 2015, S. 81f.) sind ungeklärt und können einige davon sein. Diese werden in den folgenden Abschnitten kurz diskutiert, mit dem Hinweis darauf, dass diese Vermutungen mit Vorsicht interpretiert werden und weitere Forschung erforderlich ist, um die tatsächlichen Ursachen für die geringeren Unterschiede zwischen den Textproduktionen in Englisch und Deutsch im Vergleich zu Französisch zu identifizieren.

Englisch und Deutsch sind beides germanische Sprachen und weisen somit mehr gemeinsame Merkmale miteinander auf als mit der romanischen Sprache Französisch. Dennoch gibt es einen grossen Einfluss der französischen Sprache auf die englische Sprache, welcher sich besonders im Wortschatz an vielen Lehnwörtern aufzeigen lässt. Schätzungen zufolge stammen etwa ein Drittel des Wortschatzes der englischen Sprache aus dem Französischen (Manno, 2005, S. 171).

Wahrscheinlich sind aber auch motivationale Aspekte, welche beispielsweise bei der Untersuchung von Stöckli (2004, S. 101ff.) aufgezeigt werden, welcher die grössere Motivation für Englisch im Vergleich zu Französisch in der Deutschschweiz auf verschiedene Faktoren zurückführt. Ein wichtiger Grund dafür ist die starke Stellung der englischen Sprache als globale Lingua Franca in der heutigen Welt. Englisch wird oft als die Sprache der internationalen Kommunikation, der Wissenschaft, der Technologie und der Wirtschaft angesehen. Daher wird es als eine nützlichere Fremdsprache wahrgenommen als Französisch. Ein weiterer Faktor, der zur grösseren Motivation für das Erlernen der englischen Sprache beiträgt, ist wirtschaftlich gesehen auch die Attraktivität der englischsprachigen Länder und Kulturen, welche eine starke Anziehungskraft auf junge Menschen ausüben, wobei die englische Sprache und Kultur v.a. in der Deutschschweiz oft präsent sind und positiv wahrgenommen werden (ebd.). Des weiteren spielen auch die Schule und die Lehrerinnen und Lehrer eine wichtige Rolle bei der Motivation im Fremdsprachenunterricht. Zudem wird Englisch oft als einfacher und leichter zu erlernen angesehen als Französisch, was dazu beitragen kann, dass Schülerinnen und Schüler motivierter sind, diese Sprache zu lernen. So stellen Peyer et al. (2016b) fest, dass die intrinsische Motivation für Englisch am

Stufenübergang von der Primarschulstufe zur Sekundarstufe I stabil blieb, während sie für Französisch abnahm. Die extrinsische Motivation war in der 8H und 10H für Englisch höher als für Französisch. Während die extrinsische Motivation für Englisch stabil blieb, nahm sie für Französisch in der 10H ab (vgl. Kapitel 2.3).

Wie in Kapitel 2.2 beschrieben, sieht der Lehrplan 21 vor, dass beim Abschluss der obligatorischen Schulzeit im Sinne der Mehrsprachigkeitsdidaktik vergleichbare Kompetenzen in beiden Fremdsprachen erwartet werden (Bertschy et al., 2018, S. 3; EDK, 2011, S. 6f.; HarmoS-Konkordat, 2008, S. 3). Zwar kann anhand der Analyse kein Schluss darüber gezogen werden, ob die Grundkompetenzen am Ende der obligatorischen Schulzeit (EDK, 2011, S. 7) von den Schülerinnen und Schülern der Untersuchungsgruppe erreicht werden. Dennoch lassen Tendenzen erkennen, dass die Lernenden für die zugrunde liegende Aufgabe in der zweiten Fremdsprache Englisch bessere Ratings erhalten haben als in der ersten Fremdsprache Französisch, was in Bezug auf die lexikalische Vielfalt durch die Berechnungen des Guiraud-Index unterstützt werden kann. Die vorliegende Untersuchung lässt keine Aussagen über die Erreichung der Lehrplanziele oder das Erreichen des Niveau A2.1 im Bereich am Ende der obligatorischen Schulzeit zu, wie es bei Peyer et al. (2016b, S. 47) der Fall ist. Während bei Peyer et al. (2016b, S. 53) deutliche Unterschiede zwischen den Sprachkompetenzen in der für die BKZ-Region ersten Fremdsprache Englisch und zweiten Fremdsprache Französisch vorhanden sind, lassen sich ähnliche Tendenzen in Bezug auf die lexikalische Vielfalt und das Rating im Oberwallis erkennen, wo Französisch als erste und Englisch als zweite Fremdsprache eingeführt wird. Aufgrund der umgekehrten Einführungsreihenfolge der Fremdsprachen im deutschsprachigen Teil des Kantons Wallis, verhält sich auch die Jahreswochenlektion gegensätzlich, weshalb mehr Lektionen für die französische Sprache als für die englische Sprache aufgewendet werden. Dennoch können aufgrund der unterschiedlichen Reihenfolge bei der Einführung der Fremdsprachen oder aufgrund der Stundendotation keine Rückschlüsse gemacht werden (vgl. Peyer et al., 2016b, S. 50). Auch Manno & Cuenat (2018, S. 237) argumentieren, dass allein die Festlegung der Unterrichtsstunden nicht garantiert, dass das Ziel von HarmoS (vgl. HarmoS-Konkordat, 2008, Art. 4) erreicht wird. Die Ergebnisse der vorliegenden Arbeit bestätigen die Vermutung von Peyer et al. (2016b., S. 50), dass sich die Kompetenzen in den beiden Fremdsprachen, zumindest hinsichtlich der untersuchten Kriterien (lexikalische Vielfalt, oder auch im Rating

zur Aufgabenstellung, Grammatik und Rechtschreibung), am Ende der obligatorischen Schulzeit nicht angleichen und vergleichbar sind. Da die beiden Fremdsprachen typologisch betrachtet unterschiedlich und somit nicht direkt vergleichbar sind (Treffers-Daller et al., 2016, S. 11), werden Unterschiede zwischen den algebraisch ermittelten Massen der lexikalischen Vielfalt mit einem erhobenen Referenzkorpus zur identischen Fragestellung kontrastiert und in der Auswertung der Hypothese 6 diskutiert.

4.6 Hypothese 6

Die Hypothese besagt, dass die lexikalische Vielfalt in den französischen Textproduktionen der Untersuchungsgruppe stärker von einem gesammelten Referenzkorpus abweicht als in den englischen Textproduktionen. Die vorliegenden Daten scheinen diese Hypothese zu unterstützen. Der Guiraud-Index und die MTLD werden sowohl für die Untersuchungs- als auch für die Vergleichsgruppe in beiden Sprachen ermittelt. Die Auswertungen (vgl. Abbildung 8 und 9) zeigen, dass die Untersuchungsgruppen, welche die jeweiligen Sprachen als Fremdsprachen in der Schule lernen, insgesamt niedrigere Guiraud-Index-Werte und MTLD-Werte aufweisen als die Vergleichsgruppen, welche in der jeweiligen Sprache unterrichtet werden. Dies führt zum logischen Schluss, dass die lexikalische Vielfalt der Untersuchungsgruppe in beiden Sprachen niedriger ist. Die Untersuchung zielt aber darauf ab, zu zeigen, inwiefern sich die ermittelten Werte der Untersuchungsgruppe von den ermittelten Werten der lexikalischen Vielfalt der Vergleichsgruppen unterscheiden.

Insbesondere für die französischen Textproduktionen kann aufgezeigt werden (vgl. Abbildung 8 und 9), dass die Untersuchungsgruppe im Vergleich zur Vergleichsgruppe einen deutlich niedrigeren Guiraud-Index und einen niedrigeren Wert für die MTLD aufweist. Der Unterschied zwischen der Untersuchungs- und Vergleichsgruppe bei den französischen Aufsätzen ist grösser als in den englischen Textproduktionen. Dies unterstützt die Hypothese, dass die lexikalische Vielfalt in den französischen Textproduktionen der Untersuchungsgruppe stärker von einem gesammelten Referenzkorpus abweicht als in den englischen Textproduktionen.

4.7 Hypothese 7

Die vorliegenden Daten scheinen ebenfalls eine Bestätigung für die aufgestellte Hypothese zu liefern, dass Schülerinnen und Schüler mit einer hohen lexikalischen Vielfalt in Deutsch auch in der ersten und zweiten Fremdsprache eine hohe lexikalische Vielfalt aufweisen (vgl. Abbildung 10). Die Korrelationskoeffizienten zwischen den mittels Ratings beurteilten Texten in Deutsch, Französisch und Englisch zeigen eine positive lineare Beziehung zwischen der lexikalischen Vielfalt in diesen Sprachen. Die höchste Korrelation besteht zwischen Deutsch und Englisch, gefolgt von Französisch und Englisch sowie Deutsch und Französisch. Diese Ergebnisse legen nahe, dass Personen, die in Deutsch über eine grössere lexikalische Vielfalt beim Schreiben verfügen, auch tendenziell mehr Wörter in Englisch und Französisch kennen. Es ist jedoch wichtig zu beachten, dass Korrelationen nur eine Beziehung zwischen den Variablen anzeigen und keine Kausalität. Das bedeutet, dass eine hohe Korrelation zwischen den Ratings der lexikalischen Vielfalt in den verschiedenen Sprachen nicht unbedingt bedeutet, dass das Wissen in einer Sprache das Wissen in einer anderen Sprache verursacht oder umgekehrt.

Schlussendlich zeigen die Ergebnisse, dass menschliche Bewertungen tatsächlich bis zu einem gewissen Grad vorhergesagt werden können, und auch, dass eine einfach anzuwendende Formel, der Guiraud-Index (Guiraud, 1954) sowohl beim vorliegenden Korpus wie auch bei Vanhove et al. (2019) relativ gut funktioniert. Es bleibt anzumerken, dass ein valides Vielfaltsmass, mit welchem Texte unterschiedlicher Länge miteinander verglichen werden können, nach wie vor nicht vorhanden ist und in der Vielfaltsforschung gewissermassen als «Heiliger Gral» bezeichnet werden kann (Treffers-Daller et al., 2016, S. 4).

5 Fazit

Ausgangspunkt dieser Untersuchung war die Frage, inwiefern sich die lexikalische Vielfalt in Texten von Lernenden am Ende der obligatorischen Schulzeit in der ersten Fremdsprache Französisch im Vergleich zur zweiten Fremdsprache Englisch im deutschsprachigen Teil des Kantons Wallis unterscheiden. Die Datenerhebung erfolgte an verschiedenen Schulen im Oberwallis, wobei in der Untersuchungsgruppe insgesamt 306 Textproduktionen in den Sprachen Deutsch, Französisch und Englisch gesammelt wurden. Zudem wurde ein Referenzkorpus aus 37 französischen und 43 englischen Textproduktionen erstellt (vgl. Tabelle 3).

Die Ergebnisse zeigen, dass die Textproduktionen in Englisch eine höhere lexikalische Vielfalt sowohl mittels algebraischer Masse als auch beim Rating durch menschliche Gutachterinnen und Gutachter aufweisen als in Französisch. Zudem korrelieren die algebraisch berechneten Masse der lexikalischen Vielfalt mit den Ergebnissen aus den Ratings, insbesondere mit dem Guiraud-Index, was bei der Untersuchung der Unterrichtssprache Deutsch, der ersten Fremdsprache Französisch und der zweiten Fremdsprache Englisch darauf hindeutet, dass ein höheres Rating der lexikalischen Vielfalt mit einem höheren Guiraud-Index einhergeht und dass sich dieser somit als Indikator für die lexikalische Vielfalt beim Schreiben, unabhängig von der Sprache, anbietet (vgl. Kapitel 3.5).

Die Analysen zeigen, dass es einen Zusammenhang zwischen der Wahrnehmung der lexikalischen Vielfalt und der Wahrnehmung in Bezug auf die Erfüllung der Aufgabenstellung, die Grammatik und die Rechtschreibung gibt. Dies deutet darauf hin, dass es einen positiven Zusammenhang zwischen der Wahrnehmung der lexikalischen Vielfalt im Bereich Schreiben und der Wahrnehmung anderer sprachlicher Fähigkeiten, welche im Lehrplan 21 und vom GER verlangt werden (vgl. Kapitel 2.1.1, 2.2), gibt. Dabei sind die Korrelationen der Ratings zwischen der lexikalischen Vielfalt in den französischsprachigen Textproduktionen am höchsten, gefolgt von Englisch und Deutsch.

Die Untersuchungen zeigen in der Gegenüberstellung der Ratings und der Notendurchschnitte der Lernenden in der jeweiligen Sprache im Bereich Schreiben, dass es eine positive Korrelation zwischen den Notendurchschnitten und dem Mittelwert der Ratings der untersuchten Texte in allen drei untersuchten Sprachen gibt. Insgesamt weisen die Ergebnisse

der Regressionsanalysen darauf hin, dass die Schulleistungen im Bereich Schreiben mit den Fähigkeiten der Schülerinnen und Schüler in Bezug auf lexikalische Vielfalt, Aufgabenstellung, Grammatik und Rechtschreibung korrelieren. Dabei gilt es festzuhalten, dass der Unterschied zwischen den Texten der Unterrichtssprache Deutsch und der zweiten Fremdsprache Englisch kleiner ist als der Unterschied zwischen den Textproduktionen beider Fremdsprachen, sowie derjenige zwischen deutschen und französischen Aufsätzen.

Da direkte Vergleiche zwischen den beiden Fremdsprachen aufgrund typologischer Unterschiede nicht möglich sind (Treffers-Daller et al., 2016, S. 11), zeigen die Vergleiche der Berechnungen der lexikalischen Vielfalt mit der jeweiligen Fremdsprache und einer Vergleichsgruppe mit entsprechender Fremdsprache als Unterrichtssprache, dass die ermittelten Werte der lexikalischen Vielfalt bei den französischen Textproduktionen stärker vom Referenzkorpus abweichen als die Masse der englischen Aufsätze (vgl. Abbildungen 8 und 9). Auffallend ist, dass die Werte für den Guiraud-Index der Untersuchungsgruppe in der englischen Sprache praktisch identisch zu sein scheinen mit den Werten der Vergleichsgruppe.

Inwiefern ein Zusammenhang zwischen der lexikalischen Vielfalt der Unterrichtssprache und der lexikalischen Vielfalt in den beiden Fremdsprachen besteht, sind anhand von Tendenzen erkennbar (vgl. Abbildung 10). Die Korrelationskoeffizienten zwischen den deutschen und englischen Aufsätzen sowie zwischen den deutschen und französischen Texten sind ähnlich hoch, gefolgt von den Korrelationen zwischen den Französisch- und Englischaufsätzen. Dies bedeutet, dass Personen, die in Deutsch über eine grössere lexikalische Vielfalt im Bereich Schreiben verfügen, tendenziell auch mehr Wörter in Englisch und Französisch kennen.

Insgesamt lässt sich sagen, dass die vorliegende Untersuchung wichtige Hinweise dazu liefert, wie sich die lexikalische Vielfalt in Texten von Schülerinnen und Schülern am Ende der obligatorischen Schulzeit in der ersten Fremdsprache Französisch im Vergleich zur zweiten Fremdsprache Englisch im deutschsprachigen Teil des Kantons Wallis unterscheidet. Die Ergebnisse zeigen, dass die Textproduktionen in der zweiten Fremdsprache Englisch tendenziell eine höhere lexikalische Vielfalt aufweisen als die Textproduktionen in der ersten Fremdsprache Französisch. Diese Ergebnisse können als eine Grundlage für die Weiterentwicklung des Fremdsprachenunterrichts dienen.

Es gibt jedoch auch einige Limitationen dieser Arbeit, die in zukünftigen Untersuchungen berücksichtigt werden sollten:

- Es wurden nur Schülerinnen und Schüler im deutschsprachigen Teil des Kantons Wallis untersucht, so dass die Ergebnisse möglicherweise nicht auf andere Regionen übertragbar sind. Zudem besteht die Vergleichsgruppe in den Konditionen Französisch und Englisch aus Lernenden von einem oder zwei Schulzentren. Ausserdem wurde nur eine relativ kleine Anzahl von Texten untersucht, so dass die Ergebnisse möglicherweise nicht repräsentativ sind. Zukünftige Studien könnten daher eine grössere Stichprobe von Texten aus verschiedenen Regionen untersuchen, um die Generalisierbarkeit der Ergebnisse zu verbessern und eine grössere Repräsentativität zu bewirken. Nichtsdestotrotz können mit der vorliegenden Arbeit aber auch Tendenzen ausgemacht werden, welche unabhängig von der Einführungsreihenfolge der Fremdsprachen bestehen und entweder Vermutungen bezüglich der Wochenstundendotation kritisch beleuchten oder bestehende Studien bestätigen (z.B. Peyer et al., 2016b).
- Es gilt zu beachten, dass der Französischunterricht in den letzten beiden Schuljahren in zwei Niveaunklassen stattfindet, währenddem der Englischunterricht in ganzen Klassenverbänden stattfindet. In der vorliegenden Untersuchung wurde keine Unterscheidung zwischen den Niveaus vorgenommen.
- In der gegenwärtigen Arbeit ist die Form der schriftlichen Textproduktionen stark an eine Aufgabenstellung gebunden. In zukünftigen Untersuchungen können auch andere Formen der schriftlichen Textproduktionen gewählt werden. Es wäre interessant, auch andere Textarten wie z.B. Essays oder Berichte zu untersuchen und zu vergleichen, ob sich dabei ähnliche Muster zeigen.
- Es könnte tatsächlich eine Rolle spielen, dass die deutschsprachigen Texte in der Muttersprache verfasst werden und daher von den angehenden Lehrpersonen höhere Erwartungen an die sprachliche Qualität gestellt werden. Auf der anderen Seite könnten die englischsprachigen Texte trotz möglicher sprachlicher Fehler und Einschränkungen aufgrund des Fremdsprachenstatus eine höhere Beurteilung erhalten, da sie als Leistung im Umgang mit der Fremdsprache angesehen werden. In Anlehnung an Vanhove et al. (2019, S. 518) müsste auch darüber nachgedacht werden, die Anzahl der Bewertungen pro Text zu erhöhen (vgl. Kapitel 3.4.2.4), um eine höhere

Objektivität zu gewährleisten und mögliche Verzerrungen durch individuelle Meinungen zu minimieren. Hier könnten beispielsweise auch Bewertungen von Gutachterinnen und Gutachtern französischer bzw. englischer Muttersprache einfließen.

- In der vorliegenden Arbeit werden lexikalische Vielfaltsmasse und Resultate aus den holistischen Ratings verglichen. Um eine differenziertere Aussage über die Erreichung der Kompetenzen im Lehrplan 21 und der erforderlichen GER Niveaus (vgl. Kapitel 2.1.1, 2.2) zu machen, sind zusätzliche Untersuchungen, beispielsweise in Form von Sprachkompetenztests, nötig.
- Diese Arbeit beschränkt sich auf die Analyse der Unterschiede zwischen der Unterrichtssprache und den beiden Fremdsprachen. Mit der Auswertung der zusätzlich erhobenen Daten mittels Fragebogen (vgl. Kapitel 8.2) können möglicherweise zusätzliche Informationen über mögliche Gründe für die Unterschiede der lexikalischen Vielfalt zwischen den beiden Fremdsprachen bezüglich der Motivation und Einstellung gesammelt werden (vgl. Kapitel 4.5).

Die vorhandenen Analysen zeigen Hinweise auf Unterschiede in der lexikalischen Vielfalt zwischen den beiden unterrichteten Fremdsprachen. Allerdings ist es schwierig, klare Schlussfolgerungen bezüglich der lexikalischen Vielfalt in Verbindung mit den im Lehrplan 21 definierten Zielen zu ziehen. Weitere Studien sind notwendig, um diese Zusammenhänge genauer zu untersuchen. Bei zukünftigen Untersuchungen sollten folgende Aspekte berücksichtigt werden:

- In weiteren Studien ist es von entscheidender Bedeutung, auch andere Kompetenzbereiche des Lehrplans 21 zu berücksichtigen. Während die vorliegenden Analysen Hinweise auf Unterschiede bezüglich der lexikalischen Vielfalt beim Schreiben zwischen den beiden unterrichteten Fremdsprachen und der Unterrichtssprache liefern, können Tendenzen in den anderen Kompetenzbereichen unentdeckt bleiben. Daher sollten zukünftige Studien neben dem Bereich Sprechen auch rezeptive Fähigkeiten untersuchen, um ein umfassendes Bild der lexikalischen Vielfalt zu erhalten. Darüber hinaus könnten zukünftige Studien auch andere Aspekte der Sprachproduktion untersuchen, wie z.B. die syntaktische und semantische Komplexität von Texten.

- Eine Langzeitstudie würde dazu beitragen, ein umfassenderes Verständnis dafür zu entwickeln, wie sich die lexikalische Vielfalt von Lernenden im Laufe ihrer Schulzeit entwickelt und wo die Schnittstellen zwischen den beiden unterrichteten Fremdsprachen liegen. Die vorliegende Studie hat nur den Endzustand der Sprachproduktion im Bereich Schreiben untersucht. Es wäre interessant zu untersuchen, wie sich die lexikalische Vielfalt während des Spracherwerbsprozesses entwickelt und welche Faktoren hier eine Rolle spielen. Dies könnte ebenfalls dazu beitragen, ein umfassenderes Verständnis dafür zu entwickeln, wie sich die lexikalische Vielfalt bei Schülerinnen und Schülern im Laufe der Schulzeit entwickelt und welche Faktoren dazu beitragen. Eine solche Langzeitstudie wäre jedoch mit erheblichem Aufwand verbunden, da die Lernenden über mehrere Jahre hinweg beobachtet werden müssten. Dennoch wäre es möglicherweise eine lohnenswerte Investition, da sie dazu beitragen würde, das Verständnis für den Spracherwerbsprozess und die Aneignung von Wortschatz zu verbessern und somit längerfristig den Sprachunterricht zu optimieren.
- Weitere Forschungsfragen können darauf abzielen, die Ursachen für die Unterschiede in der lexikalischen Vielfalt zwischen den beiden Fremdsprachen und der Unterrichtssprache genauer zu analysieren (vgl. Kapitel 4). Die vorliegende Studie hat gezeigt, dass die lexikalische Vielfalt in der zweiten Fremdsprache Englisch tendenziell höher ist als in der ersten Fremdsprache Französisch. Eine interessante Fragestellung könnte nun damit erweitert werden, Vermutungen zu überprüfen und darzulegen, welche Faktoren hierfür verantwortlich sind.
- Es lässt sich festhalten, dass die vorliegende Studie aufzeigt, dass die lexikalische Vielfalt in den beiden Fremdsprachen am Ende der obligatorischen Schulzeit möglicherweise ungleich verteilt ist, ohne dass eine Bewertung des Niveaus nach dem GER vorliegt. Dieses Ergebnis liefert wichtige Impulse für zukünftige Überlegungen zur Verbesserung der Didaktik im Fremdsprachenunterricht, um unter verschiedenen Bedingungen die angestrebten Ziele des Lehrplans erreichen zu können. Weitere Forschungsarbeiten sind notwendig, um die Gründe für die unterschiedliche lexikalische Vielfalt in den beiden Sprachen zu erforschen und geeignete Unterrichtsmethoden zu entwickeln, um den Schülerinnen und Schülern ein ausgewogenes Niveau in beiden Sprachen zu ermöglichen.

- Jarvis (2013b) weist darauf hin, dass menschliche Bewertungen des lexikalischen Reichtums bis zu einem gewissen Grad vorhersehbar sind, was diese Untersuchung in Bezug auf die Berechnung des Guiraud-Index ebenfalls aufzeigt. Es könnte fruchtbar sein, diese Beziehung ebenfalls in anderen Sprachen zu analysieren und andere Operationalisierungen der von ihm vorgeschlagenen Dimensionen zu prüfen (vgl. Kapitel 2.4.2.8).
- Insgesamt wäre es also sinnvoll, verschiedene statistische Modelle zu untersuchen und zu vergleichen, um festzustellen, welches am besten geeignet ist, die Beziehung zwischen textbasierten Metriken des lexikalischen Reichtums und der wahrgenommenen lexikalischen Vielfalt zu erklären. Es ist denkbar, dass nicht-lineare Beziehungen zwischen den Indizes und den Wahrnehmungen bestehen. In diesem Fall wäre es lohnenswert, mit statistischen Modellen zu experimentieren, die keine linearen Beziehungen zwischen den Indizes und den Wahrnehmungen annehmen (Vanhove et al., 2019, S. 503). Darüber hinaus ist es auch möglich, dass die textbasierten Indizes in nicht-additiven Beziehungen zu den Wahrnehmungen des lexikalischen Reichtums stehen, ähnlich wie Interaktionseffekte bei der Beurteilung der Schreibqualität (Jarvis et al., 2003). In diesem Fall könnte es sich lohnen, Interaktionseffekte in den Modellen zu berücksichtigen.

6 Literaturverzeichnis

- Alderson, J. C. (2007). The CEFR and the Need for More Research. *The Modern Language Journal*, 91(4), 659–663. https://doi.org/10.1111/j.1540-4781.2007.00627_4.x
- Baayen, R. H. (2001). *Word Frequency Distributions*. Springer Science & Business Media. [https://books.google.ch/books?hl=de&lr=&id=FWicxYWJP9sC&oi=fnd&pg=PR9&dq=Baayen,+R.+H.+\(2001\).+Word+frequency+distributions.+Dordrecht,+Netherlands:+Kluwer.&ots=XkIEYJInFo&sig=18JXpaHZ2ILUloPgX4xrj2dZKil#v=onepage&q&f=false](https://books.google.ch/books?hl=de&lr=&id=FWicxYWJP9sC&oi=fnd&pg=PR9&dq=Baayen,+R.+H.+(2001).+Word+frequency+distributions.+Dordrecht,+Netherlands:+Kluwer.&ots=XkIEYJInFo&sig=18JXpaHZ2ILUloPgX4xrj2dZKil#v=onepage&q&f=false)
- Berman, R. A., Nayditz, R., & Ravid, D. (2011). Linguistic diagnostics of written texts in two school-age populations. *Written Language & Literacy*, 14(2), 161–187. <https://doi.org/10.1075/wll.14.2.01ber>
- Bertschy, I., Egli Cuenat, M., & Stotz, D. (2018). *Passepartout. Fremdsprachen an der Volkshochschule. Lehrplan Französisch und Englisch*. (Version Kanton Wallis; S. 80). <https://www.vs.ch/documents/212242/1259225/Passepartout+Lehrplan+Franz%C3%B6sisch+und+Englisch.pdf/df0b8d64-e4e5-497b-bb94-6b47af9127d2>
- Bono, M. (2008). *La compétence plurilingue: Regards francophones* (D. Moore & V. Castellotti, Hrsg.). Lang.
- Bonvin, A., & Lambelet, A. (2017). Algorithmic and subjective measures of lexical diversity in bilingual written corpora: A discussion. *Corela. Cognition, représentation, langage*, HS-21, Article HS-21. <https://doi.org/10.4000/corela.4843>
- Bonvin, A., & Lambelet, A. (2019). Exploration empirique de la richesse lexicale: La perception humaine. *Linguistik Online*, 100(7), 65–94. <https://doi.org/10.13092/lo.100.6018>
- Bonvin, A., Vanhove, J., Berthele, R., & Lambelet, A. (2018). Die Entwicklung von produktiven lexikalischen Kompetenzen bei Schüler(innen) mit portugiesischem Migrationshintergrund in der Schweiz. *Zeitschrift für Interkulturellen Fremdsprachenunterricht, Didaktik und Methodik im Bereich Deutsch als Fremdsprache*(Nummer 1), 135–148.
- Carroll, J. B. (1938). Diversity of vocabulary and the harmonic series law of word-frequency distribution. *The Psychological Record*, 2, 379–386.
- Carroll, J. B. (1967). On sampling from a lognormal model of word frequency distribution. In H. Kučera & F. Nelson (Hrsg.), *Computational analysis of present-day American English* (S.

406–424). Brown University Press.

- Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2), 94–100. <https://doi.org/10.1080/09296171003643098>
- Crossley, S. A., & McNamara, D. S. (2011). *Shared features of L2 writing: Intergroup homogeneity and text classification*. 271–285.
- Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, 28(4), 561–580. <https://doi.org/10.1177/0265532210378031>
- Crossley, S., Salsbury, T., & McNamara, D. (2009). Measuring L2 Lexical Growth Using Hypernymic Relationships. *Language Learning*, 59(2), 307–334. <https://doi.org/10.1111/j.1467-9922.2009.00508.x>
- Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical Richness in the Spontaneous Speech of Bilinguals. *Applied Linguistics*, 24(2), 197–222. <https://doi.org/10.1093/applin/24.2.197>
- Daller, M. (2010). *Guiraud's index of lexical richness*. <https://uwe-repository.worktribe.com/output/975505>
- Davies, A. (2003). *The native speaker: Myth and reality* (2nd ed). Multilingual Matters.
- D-EDK. (2016). *Lehrplan 21: Französisch 1. Fremdsprache: Kompetenzaufbau 2. - 3. Zyklus*. Abgerufen 2. März 2023, von <https://v-fe.lehrplan.ch/impressum.php>
- Departement für Bildung und Sicherheit. (2016). *Interne Weisungen für die Kompetenzen in den Sprachen 2 und 3, die Austauschsemester und den bilingualen Kursus in der Bachelorausbildung für die Basis- und Primarstufe*.
- Departement für Volkswirtschaft und Bildung des Kantons Wallis. (2018). *Lehrplan 21: Gesamtausgabe. Lehrplan für die deutschsprachigen Schulen des Kantons Wallis*. https://vs.lehrplan.ch/lehrplan_printout.php?e=1&k=1
- Dörnyei, Z., & Chan, L. (2013). Motivation and Vision: An Analysis of Future L2 Self Images, Sensory Styles, and Imagery Capacity Across Two Target Languages. *Language Learning*, 63(3), 437–462. <https://doi.org/10.1111/lang.12005>
- Dörnyei, Z., & Ushioda, E. (2011). *Teaching and researching motivation* (2nd ed).

Longman/Pearson.

Duran, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental Trends in Lexical Diversity. *Applied Linguistics*, 25(2), 220–242. <https://doi.org/10.1093/applin/25.2.220>

EDK. (2011). *Grundkompetenzen für die Fremdsprachen. Nationale Bildungsstandards*.

EDK. (2022). *Erste obligatorische Fremdsprache / première langue étrangère obligatoire Primarstufe / degré primaire – Sekundarstufe I / degré secondaire I* (S. 1–2). <https://www.edk.ch/de/bildungssystem/kantonale-schulorganisation/kantonsumfrage/b-11-fremdsprachen-sprache-beginn>

EDK. (2004, März 25). *Sprachunterricht in der obligatorischen Schule: Strategie der EDK und Arbeitsplan für die Gesamtschweizerische Koordination*. BESCHLUSS VOM 25. MÄRZ 2004.

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4(2), 139–155. [https://doi.org/10.1016/1060-3743\(95\)90004-7](https://doi.org/10.1016/1060-3743(95)90004-7)

Europarat (Hrsg.). (2001). *Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, lehren, beurteilen*. Langenscheidt. <https://www.goethe.de/z/50/commeuro/i2.htm>

Fergadiotis, G., Wright, H. H., & Green, S. B. (2015). Psychometric Evaluation of Lexical Diversity Indices: Assessing Length Effects. *Journal of Speech, Language, and Hearing Research*, 58(3), 840–852. https://doi.org/10.1044/2015_JSLHR-L-14-0280

Fergadiotis, G., Wright, H. H., & West, T. M. (2013). Measuring Lexical Diversity in Narrative Discourse of People With Aphasia. *American Journal of Speech-Language Pathology*, 22(2). [https://doi.org/10.1044/1058-0360\(2013/12-0083\)](https://doi.org/10.1044/1058-0360(2013/12-0083))

Gardner Robert C. (2004). *The Attitude/Motivation Test Battery: International AMTB Research Project*. The University of Western Ontario, Canada.

Greber, L. (2018). *Textproduktion auf der Primarschulstufe*. <https://folia.unifr.ch/global/documents/307502>

Guiraud, P. (1954). *Les caractères statistiques du vocabulaire: Essai de méthodologie*. Presses universitaires de France, 1954.

HarmoS-Konkordat, SGS 411.4 § Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (2008). https://lex.vs.ch/app/de/texts_of_law/411.4

- Heinzmann, S. (2013). *Young Language Learners' Motivation and Attitudes: Longitudinal, Comparative and Explanatory Perspectives*. Bloomsbury Publishing; mzh. <http://search.ebscohost.com/login.aspx?direct=true&db=mzh&AN=2013902165&site=ehost-live>
- Herdan, G. (1964). *Quantitative linguistics*. Butterworths.
- Honoré, A. (1979). Some simple measures of richness of vocabulary. In *Association for Literary and Linguistic Computing Bulletin: Bd. 7(2)* (S. 172–177).
- Housen, A., Schoonjans, E., Janssens, S., Welcomme, A., Schoonheere, E., & Pierrard, M. (2011). Conceptualizing and measuring the impact of contextual factors in instructed SLA – the role of language prominence. *IRAL - International Review of Applied Linguistics in Language Teaching*, 49(2). <https://doi.org/10.1515/iral.2011.005>
- Hulstijn, J. H. (2007). The Shaky Ground Beneath the CEFR: Quantitative and Qualitative Dimensions of Language Proficiency. *The Modern Language Journal*, 91(4), 663–667. https://doi.org/10.1111/j.1540-4781.2007.00627_5.x
- Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing*, 19(1), 57–84. <https://doi.org/10.1191/0265532202lt220oa>
- Jarvis, S. (2017). Grounding lexical diversity in human judgments. *Language Testing*, 34(4), 537–553. <https://doi.org/10.1177/0265532217710632>
- Jarvis, S. (2013a). Capturing the Diversity in Lexical Diversity: Lexical Diversity. *Language Learning*, 63, 87–106. <https://doi.org/10.1111/j.1467-9922.2012.00739.x>
- Jarvis, S. (2013b). Defining and measuring lexical diversity. In S. Jarvis & M. Daller (Hrsg.), *Vocabulary knowledge: Human ratings and automated measures* (S. 13–43). John Benjamins.
- Jarvis, S., & Daller, M. (Hrsg.). (2013). *Vocabulary Knowledge: Human ratings and automated measures* (Bd. 47). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.47>
- Johnson, W. (1944). I. A program of research. *Psychological Monographs*, 56(2), 1–15. <https://doi.org/10.1037/h0093508>
- Koizumi, R. (2012). Relationships between text length and lexical diversity measures: Can we use short texts of less than 100 tokens? *Vocabulary Learning and Instruction*, 01(1), 60–69.

<https://doi.org/10.7820/vli.v01.1.koizumi>

Konsortium ÜGK (Hrsg.). (2019). *Überprüfung der Grundkompetenzen. Nationaler Bericht der ÜGK 2017: Sprachen 8. Schuljahr*. EDK und SRED.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Kyle, K., & Crossley, S. A. (2015). Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application. *TESOL Quarterly*, 49(4), 757–786. <https://doi.org/10.1002/tesq.194>

Lambelet, A., & Berthele, R. (2015). *Age and foreign language learning in school*. Palgrave Macmillan.

Laufer, B. (1994). The Lexical Profile of Second Language Writing: Does It Change Over Time? *RELC Journal*, 25(2), 21–33. <https://doi.org/10.1177/003368829402500202>

Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, 16(3), 307–322. <https://doi.org/10.1093/applin/16.3.307>

Lenz, Peter, Schneider, Günther, & Studer, Thomas. (2009). *Fremdsprachen. Wissenschaftlicher Kurzbericht und Kompetenzmodell* (Konsortium HarmoS Fremdsprachen, Hrsg.; Provisorische Fassung (vor Verabschiedung der Standards)).

Malvern, D. (2008). *Lexical diversity and language development: Quantification and assessment*. Palgrave Macmillan.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical Diversity and Language Development*. Palgrave Macmillan UK. <https://doi.org/10.1057/9780230511804>

Malvern, D. & Richards, B. (1997). A measure of lexical diversity. In *Evolving models of language. Papers from the annual meeting of the British Association of Applied Linguistics held at the University of Wales* (S. 58–71). Clevedon: Multilingual Matters.

Manno, G. (2005). Manno, Giuseppe (2003, korrigierte Auflage 2005): *“Spracheinstellungen und Tertiärsprachendidaktik. Das Frühenglische in der Deutschschweiz: eine Chance für den Französischunterricht zum Neuanfang?”* in: Britta Hufeisen/Gerhard Neuner (Hg.): *Mehrsprachigkeitskonzept - Tertiärsprachenlernen - Deutsch nach Englisch*, Graz & München, Council of Europe/European Center for Modern Language, 157-176 (S. 157–176).

Manno, G., & Cuenat, M. E. (2018). *Sprachen- und fächerübergreifende curriculare Ansätze im*

Fremdsprachenunterricht in der Schweiz. Curricula in zwei Bildungsregionen und Resultate aktueller empirischer Studien in der Deutschschweiz.

- McCarthy, P. M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)*. Doctoral dissertation.
- McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing*, 24(4), 459–488. <https://doi.org/10.1177/0265532207080767>
- 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. <https://doi.org/10.3758/BRM.42.2.381>
- McCarthy, P. M., & Jarvis, S. (Hrsg.). (2013). From intrinsic to extrinsic issues of lexical diversity assessment: An ecological validation study. In S. Jarvis & M. Daller (Hrsg.), *Vocabulary knowledge. Human ratings and automated measures*. (S. 47–78). Benjamins.
- Mellor, A. (2011). *Essay length, lexical diversity and automatic essay scoring. Series B*(Vol. 55, No. 2), 1–14.
- Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing*, 26(1), 45–65. <https://doi.org/10.1007/s11145-012-9392-5>
- Peyer, E., Andexlinger, M., & Kofler, K. (2016a). *Projekt Fremdsprachenevaluation BKZ – Schlussbericht zu den Befragungen der Französisch- und Englischlehrpersonen*.
- Peyer, E., Andexlinger, M., Kofler, K., Lenz, P., Université (Fribourg). Institut de plurilinguisme, & Haute école pédagogique Fribourg (Hrsg.). (2016b). *Projekt Fremdsprachenevaluation BKZ: Schlussbericht zu den Sprachkompetenztests*.
- Polio, C. (2012). How to Research Second Language Writing. Background: Overview of Second Language Writing Research. In A. Mackey & M. S. Gass (Hrsg.), *Research Methods in Second Language Acquisition: A Practical Guide* (S. 139–157). Blackwell Publishing Ltd.
- Read, J. A. S. (2000). *Assessing vocabulary*. Cambridge University Press.
- Sauro, S., & Smith, B. (2010). Investigating L2 Performance in Text Chat. *Applied Linguistics*, 31(4), 554–577. <https://doi.org/10.1093/applin/amq007>
- Sichel, H. S. (1975). On a distribution law for word frequencies. *Journal of the American Statistics*

Association, 25–34.

Staatskanzlei Kanton Wallis. (2020). Staatsratsentscheid Studentafel Primarschule.

Staatsrat Kanton Wallis. (2014.) Übergangsstudentafel für die Orientierungsschule Oberwallis (ab 2014/2015).
<https://www.vs.ch/documents/212242/1552839/Studentafel%20OS%20.pdf/64f4aeac-3c95-4315-a494-68a1e585dedd?t=.now?long>

Stöckli, G. (2004). *Motivation im Fremdsprachenunterricht: Eine theoriegeleitete empirische Untersuchung in 5. und 6. Primarschulklassen mit Unterricht in Englisch und Französisch* (1. Aufl.). Sauerländer.

Templin, M. (1957). *Certain Language Skills in Children: Their Development and Interrelationships* (Institute of Child Welfare Monograph 26). University of Minnesota Press.

Treffers-Daller, J. (2013). Measuring lexical diversity among L2 learners of French: An exploration of the validity of D, MTLD and HD-D as measures of language ability. In S. Jarvis & M. Daller (Hrsg.), *Vocabulary knowledge: Human ratings and automated measures* (S. 79–103). John Benjamins.

Treffers-Daller, J., Parslow, P., & Williams, S. (2016). Back to Basics: How Measures of Lexical Diversity Can Help Discriminate between CEFR Levels. *Applied Linguistics*, amw009.
<https://doi.org/10.1093/applin/amw009>

Tweedie, F. J., & Baayen, R. H. (1998). How Variable May a Constant be? Measures of Lexical Richness in Perspective. *Kluwer Academic Publishers*, 32, 323–352.

Vanhove, J., Bonvin, A., Lambelet, A., & Berthele, R. (2019). Predicting perceptions of the lexical richness of short French, German, and Portuguese texts using text-based indices. *Journal of Writing Research*, 10(3), Article 3. <https://doi.org/10.17239/jowr-2019.10.03.04>

Yu, G. (2010). Lexical Diversity in Writing and Speaking Task Performances. *Applied Linguistics*, 31(2), 236–259. <https://doi.org/10.1093/applin/amp024>

Yule, G. U. (1944). The Statistical Study of Literary Vocabulary. *Journal of the Royal Statistical Society*, 107(2), 129. <https://doi.org/10.2307/2981280>

Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology*. Houghton Mifflin Company. <http://hdl.handle.net/2027/mdp.39015008729983>

Zipf, G. K. (1937). Observations of the Possible Effect of Mental Age Upon the Frequency-Distribution of Words, from the Viewpoint of Dynamic Philology. *The Journal of Psychology*, 4(1), 239–244. <https://doi.org/10.1080/00223980.1937.9917534>

7a Abbildungsverzeichnis

Abbildung 1: Berechnungen der lexikalischen Vielfalt der Untersuchungsgruppe mit algebraischen Massen	46
Abbildung 2: Rating der lexikalischen Vielfalt	48
Abbildung 3: Korrelation zwischen Rating der lexikalischen Vielfalt und Guiraud-Index nach Sprache	49
Abbildung 4: Korrelation zwischen Rating der lexikalischen Vielfalt und MTLD	50
Abbildung 5: Korrelation Rating der lexikalischen Vielfalt mit Aufgabenstellung, Grammatik und Rechtschreibung	54
Abbildung 6: Korrelation Rating der lexikalischen Vielfalt und Schulnoten	55
Abbildung 7: Effect plots	57
Abbildung 8: Vergleich Guiraud-Index Untersuchungsgruppe vs. Vergleichsgruppe	58
Abbildung 9: Vergleich MTLD Untersuchungsgruppe vs. Vergleichsgruppe	59
Abbildung 10: Korrelationsdiagramme Zusammenhang lexikalische Vielfalt zwischen den Sprachen	61

7b Tabellenverzeichnis

Tabelle 1: Counterbalancing bei der Datenerhebung	32
Tabelle 2: Textproduktionen nach Sprache und Modus	34
Tabelle 3: Anzahl Textproduktionen nach Vergleichsgruppe und Modus	34
Tabelle 4: Soziodemografische Daten (N=183)	35
Tabelle 5: Analysierte Textproduktionen nach Sprache, Modus und Gruppe	37
Tabelle 6: Probleme und Massnahmen bei der Textbearbeitung	37
Tabelle 7: Deskriptive Analyse des Textkorpus in Bezug auf Tokens*	39
Tabelle 8: Daten zum Rater Sample (N=18)	40
Tabelle 9: Korrelationen zwischen lemmatisiertem Originaltext und bearbeiteten Versionen	46
Tabelle 10: Regressionsmodell Rating vs. algebraische Masse lexikalischer Vielfalt	51
Tabelle 11: Regressionsanalyse Rating vs. Schulnoten nach Sprachen	56

8 Anhang

8.1 Aufgabenstellung Schülerinnen und Schüler

8.1.1 Pilotphase zur Erhebung der Texte

Lieber Schüler

Liebe Schülerin

In dieser Aufgabe wirst du einen kurzen Ausschnitt aus der «Sendung mit der Maus» sehen. Deine Aufgabe ist es, den Inhalt des Videos für einen Austauschschüler aus dem Unterwallis in einem kurzen Text auf Französisch zusammenzufassen. Die Übung ist Teil eines Forschungsprojektes der Universität Freiburg.

Hier einige wichtige Informationen:

- Gehe dabei möglichst detailliert auf die einzelnen Handlungen ein und beschreibe genau, was du gesehen hast.
- Beim Schreiben des Textes darfst du weder einen Diktionär noch einen Online-Übersetzer brauchen. Du darfst nur französische Wörter verwenden. Wenn du ein Wort nicht weisst, versuchst du es ebenfalls auf Französisch zu umschreiben.
- Achte auf eine saubere Darstellung und eine leserliche Schrift.
- Du wirst den Filmausschnitt einmal sehen. Beim Schauen des Films darfst du im folgenden Kasten Notizen nehmen.

Meine Stichworte:

Nun hast du 30 Minuten Zeit, deine Beschreibung zu verfassen.

Mein Text:



8.1.2 Beispiel Aufgabenstellung zur Erhebung der Texte in Englisch



Teil III - Englisch

1

Schreibe deinen persönlichen Code in den folgenden Kasten. Der Code besteht aus einem Grossbuchstaben und aus zwei Ziffern. Es ist wichtig, dass du den Code genau aufschreibst. *

Ihre Antwort eingeben

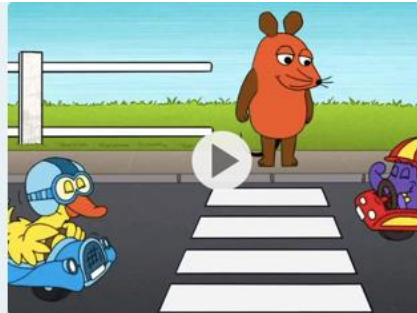
2

Selbständigkeitserklärung

Die Selbständigkeitserklärung *muß* angekreuzt werden, damit du zur Aufgabe gelangen kannst. *

- ☐ Ich bestätige hiermit, dass ich in der folgenden Aufgabe keine Hilfsmittel (Übersetzer, Diktionär, etc.) verwenden werde und dass ich die Aufgabe selbständig löse.

Englisch



Lieber Schüler
Liebe Schülerin

In dieser Aufgabe hast du einen kurzen Ausschnitt aus der «Sendung mit der Maus» gesehen. Deine Aufgabe ist es, den Inhalt des Videos für eine Schülerin aus Amerika, die daran interessiert ist zu sehen, was im Videospot zu sehen war, in einem kurzen Text auf **Englisch** wiederzugeben. Gehe dabei möglichst detailliert auf die einzelnen Handlungen ein und beschreibe genau, was du gesehen hast. Die Übung ist Teil eines Forschungsprojektes der Universität Freiburg.

Hier einige wichtige Informationen:

Während dem Film hast du Notizen gemacht. Diese helfen dir nun, das was du gesehen hast, auf Englisch aufzuschreiben.

Gehe dabei möglichst detailliert auf die einzelnen Handlungen ein und beschreibe genau, was du gesehen hast.

Beim Schreiben des Textes darfst du weder einen Diktionär noch einen Online-Übersetzer brauchen.

Du darfst nur englische Wörter verwenden. Wenn du ein Wort nicht weisst, versuchst du es ebenfalls auf Englisch zu umschreiben.

3

Schreibe deinen Text für die Schülerin aus Amerika auf **Englisch** in den folgenden Kasten: *

Ihre Antwort eingeben

8.2 Fragebogen Schülerinnen und Schüler

Liebe Schülerin

Lieber Schüler

Bis Ende des Schuljahres wirst du an einem Forschungsprojekt der Uni Fribourg mitmachen. Beim Projekt geht es darum zu untersuchen, wie du Texte auf Deutsch, Französisch und Englisch schreiben kannst. Dazu wirst du zuerst einen Fragebogen ausfüllen. Bis zum Ende des Schuljahres wirst du dreimal einen kurzen Filmausschnitt schauen. Deine Aufgabe wird es sein, das was du gesehen hast, innerhalb von 20 Minuten einmal auf Deutsch, einmal auf Französisch und einmal auf Englisch aufzuschreiben. Hilfsmittel sind nicht erlaubt.

Mit diesem Fragebogen wirst du einige Fragen über deine Einstellung zu den Sprachen Französisch und Englisch ausfüllen. Für das Ausfüllen des Fragebogens brauchst du zirka 15-20 Minuten.

Bitte lies die Fragen genau durch. Es gibt keine richtigen oder falschen Aussagen. Entscheide dich, welche Antwort für dich am ehesten zutrifft.

Die Angaben sind anonym und werden vertraulich behandelt. Du wirst dennoch darum gebeten, einige Angaben zu dir zu machen. Es lassen sich aber keine Rückschlüsse auf deine Person machen.

Herzlichen Dank für das Ausfüllen des Fragebogens!

1. Wie lautet dein Code?
2. In welchem Jahr bist du geboren? 2004 / 2005 / 2006 / 2007 / 2008
3. Geschlecht: männlich / weiblich
4. Welches ist deine Muttersprache? Deutsch, Französisch, Englisch, Italienisch, Portugiesisch, Albanisch, Bosnisch, Serbisch, Andere

Wie gut kannst du einen Text schreiben? Jetzt bist du einmal dran, deine Fähigkeiten zu beurteilen. Stell dir folgendes vor: Du schaust dir einen kurzen Filmausschnitt an und musst danach einen Aufsatz über den Inhalt des Films schreiben.

5. Du schreibst diesen Aufsatz auf Deutsch und zwar für einen Mitschüler, der leider Zuhause bleiben muss, weil er krank ist. Was denkst du, welche Note (nur ganze Noten möglich) zwischen 1 und 6 würdest du dir für diesen Aufsatz geben? (1=schlechteste Note, 6=beste Note)
6. Nun fasst du den Inhalt auch auf Englisch zusammen. Ihr habt nämlich auch eine neue Schülerin aus Amerika in der Klasse. Deine Aufgabe ist es, für die Schülerin den Inhalt des Videos in einem Text aufzuschreiben. Welche Note (nur ganze Noten möglich) würdest du dir für deinen englischen Aufsatz geben? (1=schlechteste Note, 6=beste Note)
7. Tatsächlich macht deine Klasse auch einen Briefaustausch mit einer Klasse aus dem Unterwallis. In diesem Brief sollst du auf Französisch aufschreiben, was du im Film gesehen hast. Was denkst du, welche Note (nur ganze Noten möglich) würdest du dir für deinen Text geben? (1=schlechteste Note, 6=beste Note)
8. In welcher Sprache kennst du mehr Wörter? Englisch / Französisch
9. Wie beliebt ist bei dir die Französische Sprache? 1 Stern = nicht beliebt, 6 Sterne = sehr beliebt
10. Wie beliebt ist bei dir die Englische Sprache? 1 Stern = nicht beliebt, 6 Sterne = sehr beliebt

Französisch – Einschätzung der schriftlichen Kompetenzen

11. Kreuze an, inwiefern die Aussagen auf dich zutreffen (stimmt gar nicht, stimmt eher nicht, stimmt eher, stimmt voll und ganz)
 - a. Ich kann einfache Geschichten auf Französisch aufschreiben.
 - b. Ich kann mir auf Französisch Notizen machen und kurze Mitteilungen verfassen (z.B. eine Einladung oder eine Wegbeschreibung)
 - c. Ich kann mit einfachen Sätzen und Ausdrücken über vertraute Themen auf Französisch schreiben (z.B. Tagesablauf, Aktivität, etc.)
 - d. Ich kann mit einfachen Sätzen kurze Geschichten auf Französisch aufschreiben (z.B. Bildergeschichte)
 - e. Ich kann in einfachen Sätzen auf Französisch darüber berichten, was wann und wo geschehen ist (z.B. Erlebnis)
 - f. Ich kann Gleichaltrige auf einfache Weise schriftlich über mich selbst oder alltägliche Dinge informieren (z.B. Lebenslauf, Zeitung, etc.)

- g. Ich kann einfache Texte über vertraute Themen schriftlich auf Französisch zusammenfassen.
- h. Ich kann auf Französisch in einfachen Texten über ein Ereignis berichten (z.B. Experiment, Reise, etc.)
- i. Ich kann vertraute Themen mit kurzen Texten auf Französisch und Bildern darstellen (z.B. Poster)

Englisch – Einschätzung der schriftlichen Kompetenzen

12. Kreuze an, inwiefern die Aussagen auf dich zutreffen. (stimmt gar nicht, stimmt eher nicht, stimmt eher, stimmt voll und ganz)

- a. Ich kann einfache Geschichten auf Englisch aufschreiben.
- b. Ich kann mir auf Englisch Notizen machen und kurze Mitteilungen verfassen (z.B. eine Einladung oder eine Wegbeschreibung)
- c. Ich kann mit einfachen Sätzen und Ausdrücken über vertraute Themen auf Englisch schreiben (z.B. Tagesablauf, Aktivität, etc.)
- d. Ich kann mit einfachen Sätzen kurze Geschichten auf Englisch aufschreiben (z.B. Bildergeschichte)
- e. Ich kann in einfachen Sätzen auf Englisch darüber berichten, was wann und wo geschehen ist (z.B. Erlebnis)
- f. Ich kann Gleichaltrige auf einfache Weise schriftlich über mich selbst oder alltägliche Dinge informieren (z.B. Lebenslauf, Zeitung, etc.)²
- g. Ich kann einfache Texte über vertraute Themen schriftlich auf Englisch zusammenfassen.
- h. Ich kann auf Englisch in einfachen Texten über ein Ereignis berichten (z.B. Experiment, Reise, etc.)
- i. Ich kann vertraute Themen mit kurzen Texten auf Englisch und Bildern darstellen (z.B. Poster)

Selbsteinschätzung Französisch

13. Wie gerne magst du das Fach Französisch? Wenn du einen Stern anklickst, heisst das, dass du das Fach Französisch nicht gerne magst. Wenn du alle sechs Sterne anklickst, heisst das, dass du das Fach Französisch sehr gerne magst.

14. Wie schätzt du dich im Fach Französisch ein? Wenn du einen Stern anklickst, heisst das, dass du im Fach Französisch nicht gut bist. Wenn du alle sechs Sterne anklickst, heisst das, dass du im Fach Französisch sehr gut bist.

15. Wie gross ist deine Motivation Französisch zu lernen? Wenn du die Zahl 1 wählst, heisst das, dass du keine Motivation hast Französisch zu lernen. Wenn du die Zahl 6 wählst, heisst das, dass deine Motivation Französisch zu lernen sehr gross ist.

Selbsteinschätzung Englisch

16. Wie gerne magst du das Fach Englisch? Wenn du einen Stern anklickst, heisst das, dass du das Fach Englisch nicht gerne magst. Wenn du alle sechs Sterne anklickst, heisst das, dass du das Fach Englisch sehr gerne magst.

17. Wie schätzt du dich im Fach Englisch ein? Wenn du einen Stern anklickst, heisst das, dass du im Fach Englisch nicht gut bist. Wenn du alle sechs Sterne anklickst, heisst das, dass du im Fach Englisch sehr gut bist.

18. Wie gross ist deine Motivation Englisch zu lernen? Wenn du die Zahl 1 wählst, heisst das, dass du keine Motivation hast Englisch zu lernen. Wenn du die Zahl 6 wählst, heisst das, dass deine Motivation Englisch zu lernen sehr gross ist.

Selbsteinschätzung Deutsch

19. Wie gerne magst du das Fach Deutsch? Wenn du einen Stern anklickst, heisst das, dass du das Fach Deutsch nicht gerne magst. Wenn du alle sechs Sterne anklickst, heisst das, dass du das Fach Deutsch sehr gerne magst.

20. Wie schätzt du dich im Fach Deutsch ein? Wenn du einen Stern anklickst, heisst das, dass du im Fach Deutsch nicht gut bist. Wenn du alle sechs Sterne anklickst, heisst das, dass du im Fach Deutsch sehr gut bist.

Motivation und Einstellung Französisch

21. Ich lerne Französisch... (stimmt gar nicht, stimmt eher nicht, stimmt ziemlich, stimmt genau)

- a. ...weil ich gerne etwas auf Französisch sage.
- b. ...weil ich die Sprache gerne höre.
- c. ...weil mir der Französischunterricht Spass macht.
- d. ...damit ich französische Texte verstehe (z.B. Plakate, Werbungen, Liedtexte, Zeitungsartikel, etc.)
... damit ich später Aussichten auf einen guten Job habe.

- e. ...weil viele Menschen auf der Welt Französisch sprechen.
- f. ...weil ich mich damit überall auf der Welt verständigen kann.
- g. ...damit ich Menschen aus verschiedenen Ländern kennenlernen kann.
- h. ...weil Französisch auch eine Landessprache ist.
- i. ...weil das Wallis ein zweisprachiger Kanton ist und Französisch die zweite Sprache ist.
- j. ...damit ich mich mit französischsprachigen SchweizerInnen unterhalten kann.
- k. ...damit ich mich mit Leuten aus dem Unterwallis unterhalten kann.

22. Wie stehst du zu diesen Aussagen? (stimmt gar nicht, stimmt eher nicht, stimmt ziemlich, stimmt genau)

- a. Ich lerne Französisch, weil ich muss.
- b. Ich habe Angst davor, im Französisch Fehler zu machen.
- c. Ich sage lieber nichts auf Französisch, damit ich keine falsche Antwort gebe.
- d. Ich bin oft gestresst im Französisch, weil alles so schwierig ist.
- e. Ich bin immer froh, wenn ich nichts auf Französisch sagen muss.
- f. Ich finde es wichtig, Französisch zu lernen.
- g. Später einmal, werde ich sehr gut Französisch sprechen können.
- h. Es fällt mir leicht, Französisch zu lernen.
- i. Französisch ist heutzutage wichtig.
- j. Ich lerne nur so viel Französisch, wie ich muss.
- k. Ich kann mir vorstellen, dass ich eines Tages im Ausland lebe und Diskussionen auf Französisch führe.
- l. Immer wenn ich an meine berufliche Zukunft denke, stelle ich mir vor, dass ich Französisch benutze.
- m. Ich kann mir eine Situation vorstellen, in der ich mich auf Französisch verständigen kann.
- n. Ich kann mir vorstellen, dass ich mit (inter-)nationalen Freunden oder Kollegen Französisch spreche.
- o. Ich kann mir vorstellen, dass ich Französisch spreche, als wäre Französisch meine Muttersprache.
- p. Die Dinge, die ich in Zukunft tun möchte, erfordern, dass ich Französisch kann.

23. Ich stelle mir vor, dass... (stimmt gar nicht, stimmt eher nicht, stimmt ziemlich, stimmt genau)

- a. ...Frankreich ein tolles Land ist.
- b.die französischsprachige Schweiz toll ist.
- c. ...das Unterwallis toll ist.
- d. ...Französinen und Franzosen nett sind.
- e. ...französischsprachige Schweizerinnen nett sind.
- f. ...Unterwalliserinnen und Unterwalliser nett sind.

Motivation und Einstellung Englisch

24. Ich lerne Englisch... (stimmt gar nicht, stimmt eher nicht, stimmt ziemlich, stimmt genau)

- a. ...weil ich gerne etwas auf Englisch sage.
- b. ...weil ich die Sprache gerne höre.
- c. ...weil mir der Englischunterricht Spass macht.
- d. ...damit ich englische Texte verstehe (z.B. Plakate, Werbungen, Liedtexte, Zeitungsartikel, etc.)
- e. ... damit ich später Aussichten auf einen guten Job habe.
- f. ...weil viele Menschen auf der Welt Englisch sprechen.
- g. ...weil ich mich damit überall auf der Welt verständigen kann.
- h. ...damit ich Menschen aus verschiedenen Ländern kennenlernen kann.
- i. ...weil Englisch eine internationale Sprache ist.
- j. ...weil fast alle sich auf Englisch verständigen können.
- k. ...damit ich mich mit englischsprachigen Menschen unterhalten kann.

25. Wie stehst du zu diesen Aussagen? (stimmt gar nicht, stimmt eher nicht, stimmt ziemlich, stimmt genau)

- a. Ich lerne Englisch, weil ich muss.
- b. Ich habe Angst davor, im Englisch Fehler zu machen.
- c. Ich sage lieber nichts auf Englisch, damit ich keine falsche Antwort gebe.
- d. Ich bin oft gestresst im Englisch, weil alles so schwierig ist.
- e. Ich bin immer froh, wenn ich nichts auf Englisch sagen muss.
- f. Später einmal, werde ich sehr gut Englisch sprechen können.
- g. Es fällt mir leicht, Englisch zu lernen.

- h. Englisch ist heutzutage wichtig.
 - i. Ich lerne nur so viel Englisch, wie ich muss.
 - j. Ich kann mir vorstellen, dass ich eines Tages im Ausland lebe und Diskussionen auf Englisch führe.
 - k. Immer wenn ich an meine berufliche Zukunft denke, stelle ich mir vor, dass ich Englisch benutze.
 - l. Ich kann mir eine Situation vorstellen, in der ich mich auf Englisch verständigen kann.
 - m. Ich kann mir vorstellen, dass ich mit (inter-)nationalen Freunden oder Kollegen Englisch spreche.
 - n. Ich kann mir vorstellen, dass ich Englisch spreche, als wäre Französisch meine Muttersprache.
 - o. Die Dinge, die ich in Zukunft tun möchte, erfordern, dass ich Englisch kann.
26. Ich stelle mir vor, dass... (stimmt gar nicht, stimmt eher nicht, stimmt ziemlich, stimmt genau)
- a. ...Länder, in denen Englisch gesprochen wird, tolle Länder sind.
 - b. ...die USA toll sind.
 - c. ...Grossbritannien toll ist. .
 - d. ...englischsprechende Leute nett sind.
 - e. ...Leute aus Grossbritannien nett sind.
 - f. ...Leute aus den USA nett sind.

8.3 Übersicht Lehrperson

Übersicht über die Datenerhebung – Forschungsprojekt

Schulort:

Name der Lehrperson:

Name / Vorname	Code	Frage- bogen	Zuteilung der Sprache für die Bearbeitung der Videos <i>D = Deutsch, F = Französisch, E = Englisch</i>			Niveau im Fach <i>I oder II / AP</i>		Jahresdurchschnitt im Bereich Schreiben			In Immersion
			Teil I	Teil II	Teil III	Deutsch	Franz- ösisch	Deutsch	Franz- ösisch	Englisch	
	S13	<input type="checkbox"/>	D	F	E						<input type="checkbox"/>
	S14	<input type="checkbox"/>	D	E	F						<input type="checkbox"/>
	S15	<input type="checkbox"/>	F	D	E						<input type="checkbox"/>
	S16	<input type="checkbox"/>	F	E	D						<input type="checkbox"/>

8.5. Videoclips



Die Sendung mit der Maus
Teil III



Die Sendung mit der Maus
Teil II



Die Sendung mit der Maus
Teil I

Teil I:

<https://www.youtube.com/watch?v=IJnQChnvIT4> (03.05.2022),

<https://tube.switch.ch/videos/iK344tiTjO> (03.05.2022)

Teil II:

<https://www.youtube.com/watch?v=aeccMDIeSCM> (03.05.2022),

<https://tube.switch.ch/videos/FQAQI38U43> (03.05.2022)

Teil III:

<https://www.youtube.com/watch?v=fytHXOIj7b8> (03.05.2022),

<https://tube.switch.ch/videos/OhvH6MYhjL> (03.05.2022)

8.6 Berechnungen Abbildung 1

Guiraud					
	Ymin	Q1	Median	Q3	Ymax
Deutsch	3.24	4.51	5.22	5.92	8.14
Englisch	3.27	4.23	4.76	5.48	8.16
Französisch	2.47	3.26	3.96	4.49	6.44

MTLD					
	Ymin	Q1	Median	Q3	Ymax
Deutsch	14.09	23.39	27.3	32.14	53.60
Englisch	13.48	23.55	28.48	36.95	60.16
Französisch	8.05	16.34	19.81	25.35	42.35

TTR					
	Ymin	Q1	Median	Q3	Ymax
Deutsch	0.30	0.41	0.48	0.53	0.71
Englisch	0.30	0.42	0.49	0.57	0.87
Französisch	0.26	0.45	0.50	0.58	0.88

8.7 Berechnungen Abbildung 2

	Ymin	Q1	Median	Q3	Ymax
Deutsch	1.20	2.75	3.40	4.20	4.83
Englisch	1.40	2.58	3.37	3.88	5.00
Französisch	1.00	1.73	2.20	2.81	4.25

8.8 Berechnungen Abbildung 8

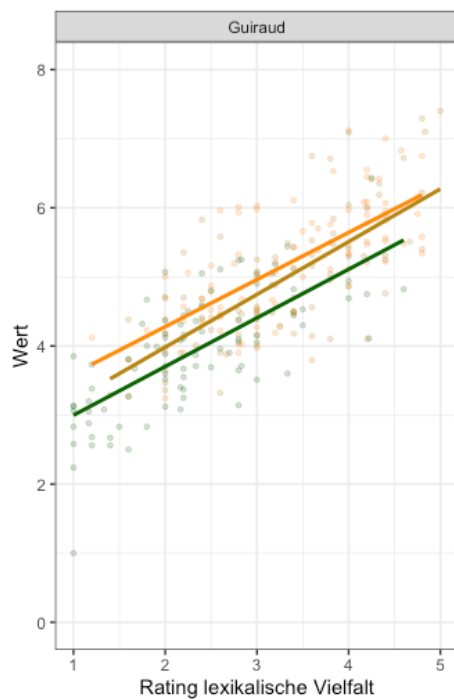
Guiraud				
Sprache	Gruppe	Median	Q1	Q3
Englisch	Vergleichsgruppe	4.82	4.15	5.59
Englisch	Untersuchungsgruppe	4.76	4.23	5.48
Französisch	Vergleichsgruppe	5	4.8	5.66
Französisch	Untersuchungsgruppe	3.96	3.26	4.49

8.9 Berechnungen Abbildung 9

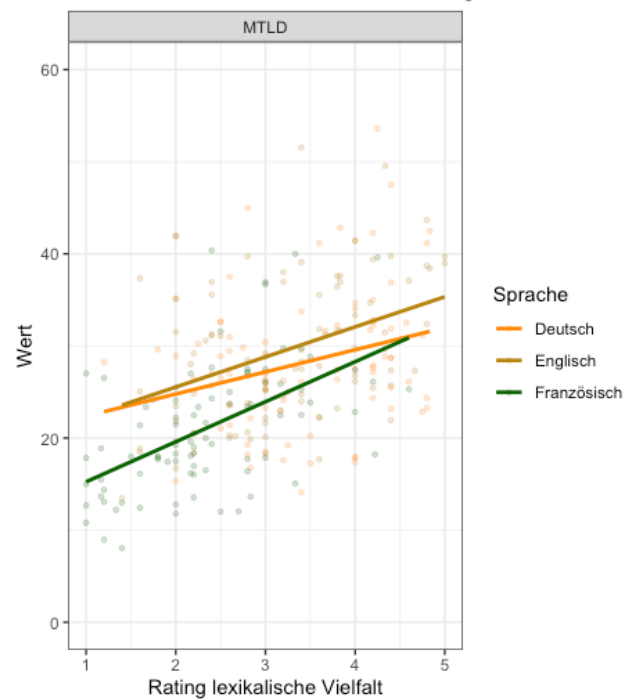
Sprache	Gruppe	MTLD		
		Median	Q1	Q3
Englisch	Vergleichsgruppe	34.465	28.9650	39.9350
Englisch	Untersuchungsgruppe	28.480	23.55	36.95
Französisch	Vergleichsgruppe	31.840	26.13	36.07
Französisch	Untersuchungsgruppe	19.81	16.3425	25.3525

8.10 Lexikalische Vielfalt: Guiraud/MTLD vs. Rating

Lexikalische Vielfalt: Guiraud vs. Rating



Lexikalische Vielfalt: MTLD vs. Rating



9 Eidesstattliche Erklärung

Ich erkläre ehrenwörtlich, dass ich meine Masterarbeit selbständig und ohne unerlaubte fremde Hilfe verfasst habe

Gampel, 23.06.2023