

ORTHOGRAFISCHE FEHLER IN DULKO- UND FALKO-TEXTEN

Orsolya Rauzs
Universität Szeged

Abstract

Der hier vorgelegte Bericht stellt das Dulko-Korpus in den Mittelpunkt, das deutschsprachige Lernertexte von ungarischen Studierenden mit guten Deutschkenntnissen (B2 / C1) enthält. Mithilfe statistischer Analysen wird in erster Linie der Frage nachgegangen, was für orthografische Fehler mit welcher Häufigkeit im Korpus vorkommen. Um zu erfahren, ob und wie die orthografische Kompetenz der ungarischen Studierenden von der Kompetenz von Deutschen abweicht, werden auch ausgewählte Texte des Falko-L1-Essaykorpus untersucht. Daneben wird angesprochen, inwiefern die Ergebnisse der Analyse des Dulko-Korpus in Bezug auf ungarische Deutschlernende als repräsentativ betrachtet werden können.

Keywords: Lernertext; orthografischer Fehler; Korpus; Statistik; Dulko; Falko

Abstract

The present paper focuses on the Dulko corpus, which contains German learner texts written by Hungarian students with good German language skills (B2 / C1). The main question being investigated is what orthographic errors occur in the corpus and with what frequency. Selected texts from the Falko L1 essay corpus will also be analysed to find out whether and how the orthographic competence of the students in the Dulko project differs from that of native speakers. It will also be discussed which results of the analysis of the Dulko corpus can be considered representative with regard to Hungarian learners of German.

Keywords: learner text; orthographic error; corpus; statistics; Dulko; Falko

1. Einleitung

Die Sprachkompetenz von ungarischen Deutschlernenden ist ein relativ wenig untersuchtes Thema. Es gibt auf diesem Gebiet zwei grundlegende korpusbasierte Arbeiten: Die longitudinale Untersuchung von Fekete (2016) stellt die Kompetenzentwicklung von gymnasialen Lernenden während einer Erhebungszeit von drei Jahren anhand von 90 Aufsätzen dar, und zwar hinsichtlich Text- und Segmentlänge, Wortschatzwachstum sowie grammatischer Komplexität und Grammatikalität in mehreren sprachlichen Bereichen. Ergänzend führt sie auch Fehleranalysen durch. Was die Deutschkenntnisse von Studierenden betrifft, liegt die umfassende Arbeit von Huber (2023) vor, in der sie 149 Essays von Germanistikstudierenden der Eötvös-Loránd-Universität Budapest auswertet. Dabei analysiert sie Fehler auf den Ebenen Orthografie, Morphologie, Syntax, Lexikologie, Semantik und Pragmatik.

Die Untersuchung der deutschen Sprachkompetenz von Germanistikstudierenden war auch das Ziel des Dulko-Projekts¹, das zwischen 2017 und 2020 an der Universität Szeged lief. Im Rahmen dieses Projekts wurde ein Korpus aufgebaut, das Aufsätze und Übersetzungen von ungarischen Studierenden der Germanistik enthält. Die Texte wurden digitalisiert und u.a. nach Fehlerkategorien annotiert. Die Auswertung der Annotation erfolgt fortlaufend in einzelnen Publikationen (vgl. Szerecz

¹ Eine ausführliche Projektbeschreibung findet sich im Dulko-Handbuch (vgl. Beeh et al. 2021).

2022; Modrián-Horváth / Kappel 2024; Molnár 2024; Schlotthauer / Trawiński 2024), zu denen auch Beiträge von Projektmitgliedern – wie auch der hier vorgelegte Bericht – gezählt werden können.

Die primäre Zielsetzung des vorliegenden Praxisberichts ist es, erste Untersuchungsergebnisse zu präsentieren, die einen Überblick über die orthografischen Fehler im Dulko-Korpus bieten. Der Fokus liegt dabei darauf, die Häufigkeit der allgemeinen Haupttypen der orthografischen Fehler im Korpus festzustellen sowie spezifische Subtypen zu identifizieren. Außerdem wird untersucht, ob die Orthografiekompetenz der Studierenden des Dulko-Projekts von der Kompetenz muttersprachlicher Deutschsprechender abweicht. Meine Hypothese ist, dass Muttersprachlerinnen und Muttersprachler in der deutschen Orthografie (vor allem in der Wortschreibung und Groß- und Kleinschreibung) besser geübt sind, weswegen sie weniger Fehler machen. Zudem wird in der vorliegenden Arbeit der Frage nachgegangen, ob die hier vorgestellten Ergebnisse eher korpuspezifisch sind oder eventuell mit denen von Huber (2023) übereinstimmen und folglich für ungarische Deutschlernende als repräsentativ gelten können.

Anhand der genannten Zielsetzungen wird das Dulko-Gesamtkorpus wie folgt analysiert: In einem ersten Schritt werden die Fehler nach der Typologie des Dulko-Projekts statistisch erfasst. Dann erfolgt die Identifizierung von Subtypen, die innerhalb der Haupttypen häufig auftreten. Dabei werden die wichtigsten Trends – sofern möglich – kurz auch mit denen bei Huber (2023) kontrastiert. Abschließend werden die so gewonnenen Erkenntnisse mit der orthografischen Kompetenz von Muttersprachlern verglichen, zu deren Einschätzung ausgewählte Texte des Falko-Essaykorpus² ausgewertet werden.

2. Korpora und Untersuchungsmethoden

Das Dulko-Gesamtkorpus (*Deutsch-ungarisches Lernerkorpus*) enthält 21.010 Tokens in 64 Texten, die von ungarischen Germanistikstudierenden verfasst wurden. Sie hatten alle Deutschkenntnisse auf dem Sprachniveau B2 oder C1³. Die Erhebungs- und Aufbereitungsmethodik des Dulko-Korpus orientiert sich am Falko-Projekt (vgl. Beeh et al. 2021: 4, 14). Das Gesamtkorpus gliedert sich in ein Übersetzungs-Teilkorpus und ein Essay-Teilkorpus (vgl. Beeh et al. 2021: 4-5, 9):

- DulkoTranslation-v1.0: 9224 Tokens, 30 Texte, Themen der Übersetzungen:
 - Ein Schüler reist nach China;
 - Sprechen als besondere Fähigkeit der Menschen;
 - Wandern in den ungarischen Wäldern im Herbst;
 - Warum weinen Frauen?
- DulkoEssay-v1.0: 11.786 Tokens, 34 Texte, Themen der Essays:
 - Die finanzielle Entlohnung eines Menschen sollte dem Beitrag entsprechen, den er für die Gesellschaft geleistet hat;
 - Der Feminismus hat den Frauen mehr geschadet als genutzt;
 - Kriminalität zahlt sich nicht aus.

Das Falko-Korpus (*Fehlerannotiertes Lernerkorpus*) wurde von Wissenschaftlerinnen und Wissenschaftlern der Freien Universität Berlin und der Humboldt-Universität zu Berlin aus Essays sowie

² Für detaillierte Informationen über das Projekt und die Teilkorpora siehe die Projektwebseite von Falko: <https://hu-berlin.de/falko> (01.10.2024).

³ Das Sprachniveau wurde nach dem Gemeinsamen Europäischen Referenzrahmen mithilfe von C-Tests festgestellt (vgl. Beeh et al. 2021: 4, 28).

Zusammenfassungen von gymnasialen Lernenden und Studierenden zusammengestellt, die Deutsch als Mutter- oder Fremdsprache sprachen (vgl. Reznicek et al. 2012: 11-12, 23-24). Für die vorliegende Untersuchung ist das folgende Teilkorpus relevant, das Essaytexte von Muttersprachlerinnen und Muttersprachlern enthält (vgl. Reznicek et al. 2012: 32-34):

- Falko-Essaykorpus L1v2.3: 68.491 Tokens, 95 Texte, Themen der Essays:
 - Die finanzielle Entlohnung eines Menschen sollte dem Beitrag entsprechen, den er für die Gesellschaft geleistet hat;
 - Der Feminismus hat den Frauen mehr geschadet als genutzt;
 - Kriminalität zahlt sich nicht aus;
 - Die meisten Universitätsabschlüsse bereiten die Studenten nicht auf die wirkliche Welt vor. Sie sind deswegen von geringem Wert.

Im Rahmen der Untersuchung wurde einerseits das Dulko-Gesamtkorpus ausgewertet. Aus dem Falko-Essaykorpus L1v2.3 wurden andererseits nur Texte von Studierenden der Freien Universität Berlin sowie von Lernenden der Carl-Siemens-Schule Berlin ausgewählt (18.762 Tokens, 28 Texte), um in der letzten Phase der Analyse (vgl. Abschnitt 4) Textgruppen von ca. ähnlichem Umfang vergleichen zu können. Diese Auswahl zielt außerdem darauf ab, beide Probandengruppen des Falko-Essaykorpus, d.h. Studierende sowie Gymnasialschülerinnen und -schüler, gleichmäßig zu berücksichtigen. Dadurch wird eine bessere Repräsentation der allgemeinen muttersprachlichen Kompetenz der deutschen Probandinnen und Probanden angestrebt.

Bei der Analyse wurde das Suchinterface ANNIS⁴ eingesetzt, in dem die Fehler in den Dulko-Texten annotiert und mit spezifischen Suchanfragen recherchierbar sind. Anhand des Dulko-Handbuchs, das das Annotationsverfahren beschreibt, sind Fehler in diesem Fall „Abweichungen zwischen einem Teil eines Texts im System der Lernaltersprache und einer (impliziten oder expliziten) *Zielhypothese* im System der Zielsprache“ (Beeh et al. 2021: 12)⁵. Im Handbuch werden folgende vier orthografische Fehlertypen unterschieden (vgl. Beeh et al. 2021: 34-36):

- Groß- und Kleinschreibung (GKS);
- Getrennt- und Zusammenschreibung (GZS), inklusive Schreibung mit Bindestrich;
- Wortschreibung (WS), d.h. fehlende oder überflüssige Buchstaben, Verwechslung von Buchstaben;
- Zeichensetzung (ZS), d.h. falsche Setzung von Satzzeichen auf Satz- und Wortebene.

Die Belege mit orthografischen Fehlern wurden zuerst mithilfe von ANNIS nach diesen Haupttypen getrennt aufgelistet bzw. statistisch ausgewertet (vgl. Abschnitt 3.1), wonach sie auch auf häufig auftretende Muster hin geprüft wurden, um Subtypen ermitteln zu können (vgl. Abschnitt 3.2).

Im Falko-Essaykorpus sind orthografische Fehler nicht als solche markiert⁶. Im Fall der analysierten Falko-Texte war für die Belegsuche in ANNIS nur die Annotationsspur ZH1Diff hilfreich, auf der bestimmte (u.a. orthografische) Abweichungen zwischen dem jeweiligen Lernertext und der

⁴ <https://korpling.german.hu-berlin.de/annis/> (01.10.2024).

⁵ Eine Zielhypothese ist „eine korrigierte Version des Lernertextes, die [...] relativ wenig vom Lernertext abweicht“ (Beeh et al. 2021: 15); sie ist also eine konstruierte korrekte Äußerung (vgl. Reznicek et al. 2012: 38-39). Im Dulko-Projekt „werden umgangssprachliche, gesprochensprachliche sowie nationale Varianten der Standardvarietäten außerhalb Deutschlands nicht als Fehler betrachtet und demzufolge nicht annotiert. Somit wird darauf geachtet, bei der Fehlerkorrektur nur Abweichungen zu korrigieren, für deren Korrektur es kodifizierte Referenzwerke [...] bzw. eine empirische Grundlage gibt. Im Zweifelsfall wurden vergleichende Korpusrecherchen vorgenommen und Muttersprachler konsultiert“ (Beeh et al. 2021: 12).

⁶ Die aktuelle Version der Annotation des Falko-Essaykorpus wird in Hirschmann et al. (2022) vorgestellt.

Zielhypothese annotiert sind. Fehlerhafte Elemente erscheinen hier allgemein als fehlende (INS), überflüssige (DEL), veränderte (CHA), geteilte (SPLIT) oder zusammengeführte (MERGE) Tokens. Bei all diesen Tokens musste ich zuerst die orthografisch relevanten selbst identifizieren und nach den vier Dulko-Fehlertypen kategorisieren. Die Falko-Belege wurden dann einer ähnlichen statistischen Analyse unterzogen wie die Dulko-Texte (vgl. Abschnitt 4).

Bei der Auswertung der Belege blieben Fehler- oder Abweichungstags, die als Konsequenzen anderer Korrekturen angesehen werden können, unberücksichtigt. Zum Beispiel ist die folgende Veränderung der Großschreibung im Falko-Text fu074d_2006_10_L1v2.3 nicht als primärer Grundfehler aufzufassen:

[Satz 29, Lernertext: ... *warum es wiederum Menschen gibt, die weit über dem durchschnittlichen Gehaltsniveau liegen.*]

Satz 30

Lernertext: *Fragen inwieweit dies moralisch richtig ist halte ich für verlogen ...*

Zielhypothese: *Zu fragen, inwieweit dies moralisch richtig ist, halte ich für verlogen ...*

Der Lernertext im Satz 30 wurde bei der Annotation in der Zielhypothese grammatisch verändert, wobei das Substantiv *Fragen* durch das Verb *fragen* ersetzt wurde. Die Veränderung in der Groß- und Kleinschreibung des Wortes wurde mit dem Abweichungstag CHA annotiert, der aber keinen echten Lernerfehler markiert.

3. Orthografische Fehler im Dulko-Korpus

3.1 Statistische Ergebnisse hinsichtlich der Hauptfehlertypen

Zur Erstellung der folgenden Statistiken erfolgte die Fehlerzählung mithilfe von ANNIS: Mit spezifischen Suchanfragen ließen sich die Gesamtzahlen der Fehler und der betreffenden Korpustexte hinsichtlich der vier orthografischen Fehlertypen des Dulko-Projekts automatisch ermitteln. Die detaillierte Darstellung der Belege im Programm ermöglichte die Zuordnung der Fehler zu den einzelnen Korpustexten.

Im Dulko-Gesamtkorpus gibt es 450 primäre (also nicht als Folgefehler entstandene) orthografische Fehler, das entspricht ca. 2 % der Gesamttokenzahl (21.010 Tokens). Die Fehler kommen in 62 Texten vor, d.h. in 97 % der Texte.

Die Verteilung der Fehler zwischen den vier Typen kann man in Tabelle 1 sehen. Ungefähr die Hälfte der Fehler gehört zum Bereich Zeichensetzung, außerdem kommen Wortschreibungsfehler ziemlich häufig vor (34 %). Über die Hälfte der Texte weist Fehler in den Kategorien Zeichensetzung, Wortschreibung sowie Groß- und Kleinschreibung auf, wobei Zeichensetzungsfehler in fast allen Texten (89 %) vorhanden sind.

Fehlertyp	Anzahl und Anteil der Fehler (100 % = 450 Fehler)	Anzahl und Anteil der Texte mit Fehlern (100 % = 64 Korpustexte)
Zeichensetzung	221 (49 %)	57 (89 %)
Wortschreibung	154 (34 %)	44 (69 %)
Groß- und Kleinschreibung	58 (13 %)	33 (52 %)
Getrennt- und Zusammenschreibung	17 (4 %)	14 (22 %)

Tabelle 1
Orthografische Fehlertypen im Dulko-Korpus

Die Texte sind bezüglich der orthografischen Korrektheit sehr unterschiedlich, wie Tabelle 2 zeigt. In den Spalten der Streuung sieht man die Grenzwerte. Dabei fällt auf, dass im Fall der individuellen Höchstwerte die Wortschreibungsfehler überwiegen und nicht die Zeichensetzungsfehler, die bei den Durchschnittswerten auf dem ersten Platz stehen. Fehler bei der Groß- und Kleinschreibung sowie der Getrennt- und Zusammenschreibung sind sowohl aufgrund der maximalen Fehlerzahlen als auch aufgrund des Durchschnitts viel seltener⁷.

	Streuung		Durchschnitt	
	pro Text	pro 100 Tokens der Texte	pro Text	pro 100 Tokens des Korpus
Fehler insgesamt	0-33	0-11,8	7,0	2,2
Zeichensetzung	0-13	0-3,3	3,5	1,1
Wortschreibung	0-20	0-7,1	2,4	0,8
Groß- und Kleinschreibung	0-5	0-1,8	0,9	0,3
Getrennt- und Zusammenschreibung	0-2	0-0,7	0,3	0,1

Tabelle 2
Streuung und Durchschnitt der orthografischen Fehler im Dulko-Korpus

Abbildung 1 stellt die Streuung der absoluten Fehlerzahlen der 64 Texte (blau) sowie die Durchschnittswerte (orange) detailliert dar. Es ist auffällig, dass die meisten Texte bei jeder Kategorie im unteren Wertebereich erscheinen, also eher wenige Fehler enthalten.

⁷ Um bei der Errechnung der Fehlerquotienten eine einheitliche Methode verwenden zu können, wurden als Bezugsbasis die Tokenzahlen der Texte genommen, die im Projekt automatisch mit der Software EXMARaLDA (Dulko) berechnet wurden (vgl. Beeh et al. 2021: 4-11). In detaillierten und fehlertypspezifischen Analysen könnten weitere Aspekte berücksichtigt werden, z.B. bei Zeichensetzungsfehlern die Anzahl von (Teil-)Satzgrenzen.

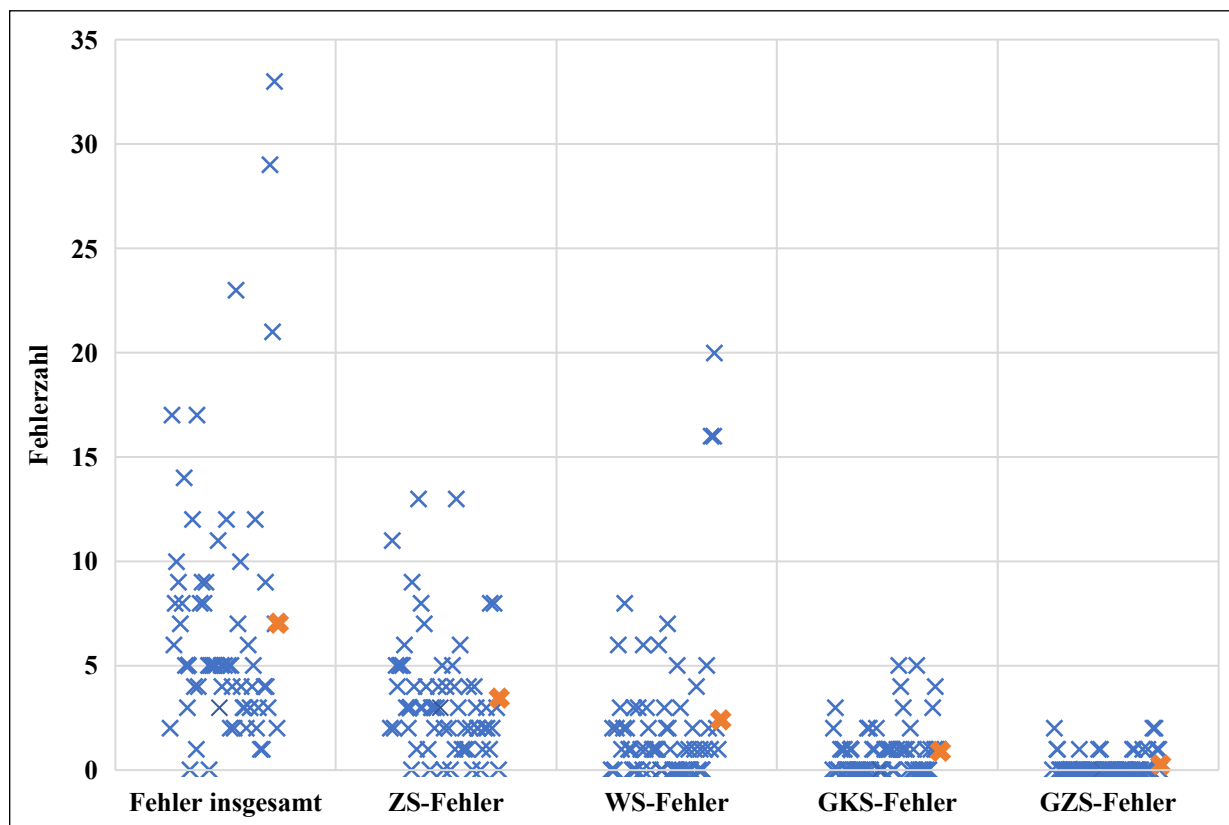


Abbildung 1
Streuung und Durchschnitt der orthografischen Fehler im Dulko-Korpus pro Text⁸

Abbildung 2 ermöglicht eine bessere Beurteilung der orthografischen Kompetenz der Studierenden, da sie die Fehlerhäufigkeit pro 100 Tokens pro Text zeigt. In Orange wurden wieder die Durchschnittswerte der Kategorien angegeben.

⁸ Für die Erklärung der Abkürzungen der Fehlerkategorien siehe Abschnitt 2.

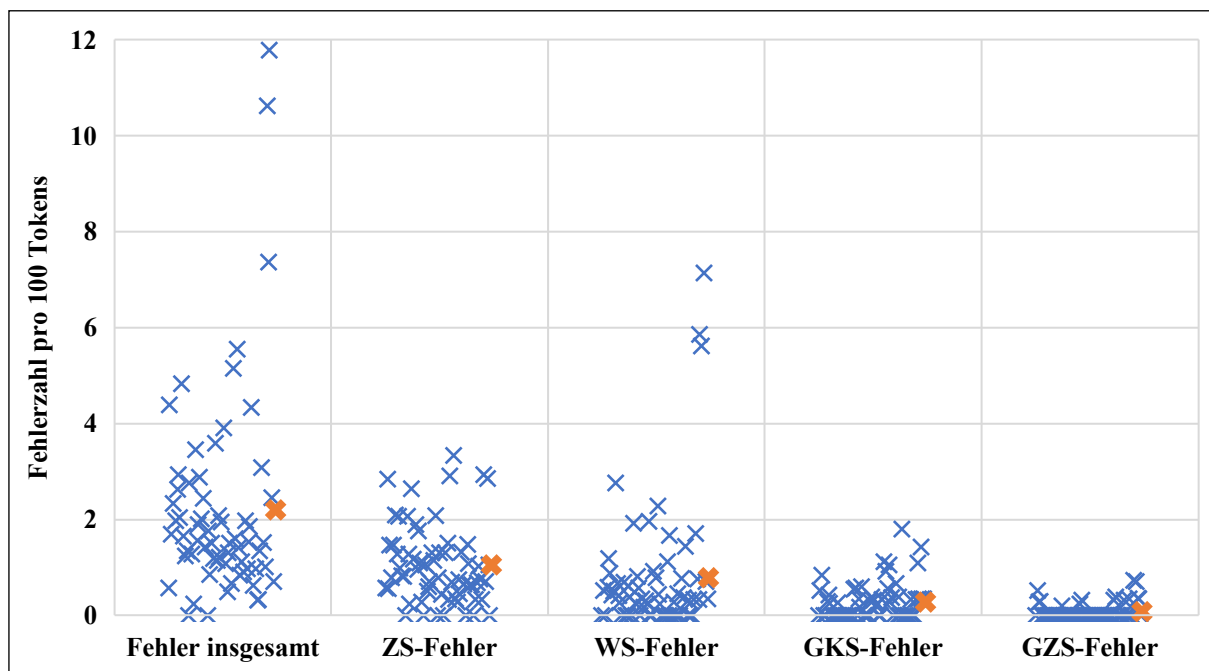


Abbildung 2
Streuung und Durchschnitt der orthografischen Fehler im Dulko-Korpus pro 100 Tokens der Texte

Was die Gesamtzahl der Fehler betrifft, bilden die Pole einerseits zwei absolut fehlerfreie Texte, andererseits ein Text mit 11,8 Fehlern pro 100 Tokens. Eine größere Streuung zwischen den minimalen und maximalen Werten findet man in der Spalte der Gesamtzahl der orthografischen Fehler und in der Spalte der Wortschreibungsfehler. In beiden Fällen sind es nur einige Studierende, die sehr viele Fehler begangen haben und weit über dem jeweiligen Durchschnitt liegen, ansonsten ist die Leistung der Studierenden relativ einheitlich, vor allem hinsichtlich Groß- und Kleinschreibung sowie Getrennt- und Zusammenschreibung.

Wenn man diese Daten mit denen von Huber (2023) vergleicht, sieht man mehrere Unterschiede⁹: Huber fand in 149 Essays 414 Fehler, die den vier Fehlertypen des Dulko-Projekts entsprechen, d.h. ihr größeres Korpus¹⁰ enthält weniger Fehler als das Dulko-Gesamtkorpus. Davon sind 33 % Zeichensetzungs-, 28 % Wortschreibungs-, 28 % Groß- und Kleinschreibungs- und 11 % Getrennt- und Zusammenschreibungsfehler (vgl. Abschnitt 4.1.2 in Huber 2023). Die Reihenfolge der Typen ist mit ihren sinkenden Anteilen zwar den Dulko-Ergebnissen ähnlich, aber die Typen selbst weisen eine ganz andere prozentuale Verteilung auf. Im Korpus von Huber (2023) sind Fehler in den Bereichen Zeichensetzung, Wortschreibung sowie Groß- und Kleinschreibung in ca. gleichem Maß vertreten, dagegen gibt es fast dreimal so viele Getrennt- und Zusammenschreibungsfehler wie in den Dulko-Texten.

⁹ Da die Fehlerklassifizierung bei Huber (2023) anders ist als im Dulko-Projekt, sind Vergleiche nur begrenzt möglich. Beispielsweise behandelt sie das Fehlen von Fugenelementen in Wörtern wie *Klasselehrerin* als Rechtschreibfehler (vgl. Huber 2023: 107), wogegen solche Fehler bei der Annotation der Dulko-Texte als Wortbildungsfehler getaggt wurden (vgl. Beeh et al. 2021: 37). Diese Beleggruppe des Huber-Korpus wird im Folgenden nicht berücksichtigt.

¹⁰ „Die Größe des Lerneraufsatzkorpus beträgt 39.049 laufende Wörter. Dabei werden Abkürzungen, Ziffern und Zeichen als Wort gezählt“ (Huber 2023: 77).

3.2 Subtypen und Beispiele

Neben der statistischen Auswertung der vier Fehlertypen des Dulko-Projekts war es ebenfalls ein Ziel der Analyse, Subtypen festzustellen. Als Grundlage diente das Amtliche Regelwerk des Rats für deutsche Rechtschreibung (2018), mit dessen Hilfe die Belege zuerst nach den Paragraphen bzw. Einzelregeln sortiert wurden. Im Fall von Belegen, bei denen eine Zuordnung nicht möglich war, musste ich die Fehlerkontexte selbst identifizieren¹¹. Danach wurden bei jedem Hauptfehlertyp aus logisch zusammenhängenden kleineren Beleggruppen größere Subtypen gebildet.

Die folgende Liste enthält Subtypen, die mindestens 10 % des jeweiligen Hauptfehlertyps ausmachen. Für jeden Subtyp werden auch Beispiele angegeben, in den meisten Fällen in Gruppen, deren Bestimmung zum größten Teil anhand des Amtlichen Regelwerks erfolgte. Die fehlerhaften Stellen werden in den Belegen mit Unterstreichungen markiert.

Zeichensetzung:

Fehlendes Komma (117 Fehler = 53 % der Zeichensetzungsfehler, 26 % aller orthografischen Fehler)

- bei Nebensatz:
... *eben wegen dieser Rolle _ die sie einfüllen müssen.* (Feminismus_7)
Das zu bestimmen, wer wie viel Geld bekommen sollte _ ist nicht einfach. (Entlohnung_16)
... wir wissen _ was Feminismus ist _ aber ... (Feminismus_3)
- bei Infinitivsatz:
Natürlich fragt man das _ um zu wissen, ... (Feminismus_20)
Ich hatte auch Lust _ mit ihnen gemeinsam turnen ... (China_1)
- bei Zusatz / Nachtrag:
nach Peking, also nach der „nördlichen Hauptstadt“ _ ... (China_4)
... wegen physischer Problemen _ z.B. wegen Hunger, Durst und Schmerzen _ weinen ... (Weinen_3)
- vor Konjunktion *aber*:
... mit Feministen über dem Thema sprechen _ aber leider ist es nicht ganz genug. (Feminismus_3)

Überflüssiges Komma (80 Fehler = 36 % der Zeichensetzungsfehler, 18 % aller orthografischen Fehler)

- vor Konjunktionen *wie / als*:
Aber in Ungarn verdienen die Menschen leider nicht so gut, wie in anderen Ländern. (Entlohnung_27)
Ich bin der Meinung, dass in die anderen Ländern die finanzielle Entlohnung besser ist, als in Ungarn. (Entlohnung_27)
- vor Konjunktion *und*:
... weil sie sehr viel arbeiten, und viel für den Gesellschaft tun. (Entlohnung_26)
... haben eine große Verantwortung, und eigentlich Macht ... (Entlohnung_9)

¹¹ Beispielsweise bei überflüssigen Kommas nach dem Vorfeld / Satzanfang (s.u.).

- Vorfeldkomma / Satzanfangskomma:
Vor Kurzem, eine Forschung wurde mit 3000 Personen gemacht. (Weinen_6)
In den meisten Länder sind, die von dem Staat angestellte Leuten gering entlohnt. (Entlohnung_10)

Im Bereich der Interpunktion dominieren Belege mit Kommasetzungsproblemen, ähnlich wie in Huber (2023). Dass die fehlende Kommasetzung auch in Hubers Analyse der bei Weitem häufigste Subtyp ist, gefolgt von der überflüssigen Kommasetzung, zeigt, dass diese Bereiche den ungarischen Deutschlernenden generell große Schwierigkeiten bereiten. Die Ursachen sind meiner Meinung nach vielfältig, z.B. fehlende syntaktische Kenntnisse (in Bezug auf Nebensatz oder Vorfeld)¹², fehlende Kenntnisse der deutschen Interpunktionsregeln (im Fall von *aber*, *und* bzw. Nachtrag), Einfluss der ungarischen Interpunktionsregeln (im Fall von *als* / *wie*) oder Flüchtigkeit. Außer der beiden häufigsten Untergruppen gibt es im Dulko-Korpus noch vereinzelte Belege mit fehlendem Anführungszeichen / Gedankenstrich / Punkt / Klammerteil oder mit einem falschen Zeichen am Satzende.

Wortschreibung:

In der folgenden Systematisierung der Belege erfolgte die Trennung des vokalischen und des konsonantischen Bereichs bzw. die Absonderung von Fremdwörtern auf der Grundlage des Amtlichen Regelwerks (2018). Es wird hier aber darauf fokussiert, welche konkreten Einzelbuchstaben falsch verwendet werden und worin das jeweilige Problem im Sinne des Dulko-Handbuchs liegt (fehlende / überflüssige Buchstaben oder Verwechslung von Buchstaben, vgl. Beeh et al. 2021: 34):

Fehler im Zusammenhang mit Konsonantenverdopplungen (35 Fehler = 23 % der Wortschreibungsfehler, 8 % aller orthografischen Fehler)

- Verdopplung fehlt:
bischen (Feminismus_12), *unvorstelbar* (Feminismus_13), *anerkant* (Feminismus_19), *männliche* (Feminismus_19)
- Verdopplung überflüssig:
offt (Weinen_12), *insgesammt* (Weinen_12), *fasst* (Weinen_14), *Überlasstungen* (Weinen_14)
- Verwechslung von *das* und *dass*:
... interessant ... ist dass, das die Dauer der Weinen ... derselbe ist. (Weinen_6)
- Probleme mit -nis:
Verhältniss (Weinen_4), *Ergebnise* (Weinen_12), *Ergebniss* (Weinen_15)

Fremdwörter (33 Fehler = 21 % der Wortschreibungsfehler, 7 % aller orthografischen Fehler)

trenierten (China_15), *Interwiew* (Feminismus_20), *Change* (Feminismus_3), *Emphat_ie* (Weinen_15)

Verwendung von einzelnen Konsonanten (22 Fehler = 14 % der Wortschreibungsfehler, 5 % aller orthografischen Fehler)

¹² Nach meinen Beobachtungen fällt es vielen Studierenden schwer, Teilsatzgrenzen in komplexen Sätzen zu identifizieren oder Hervorhebungen am Satzanfang richtig zu interpretieren.

- *t* fehlt:
Haupt_sächlich (Weinen_16), *nich_* (Sprache_8), *Letz_lich* (Weinen_5), *unterstü_zt* (Entlohnung_16)
- *t* überflüssig:
hauptsätlich (Weinen_15), *andererseitst* (Wandern_18), *Aufmerkt_samkeit* (China_15)
- *h* überflüssig:
spührt (Feminismus_9), *seehliche* (Weinen_6), *vielmahl* (Entlohnung_5), *Rheinigungs_kräfte* (Entlohnung_9)

Fehlerhafte Laut-Buchstaben-Zuordnung bei Vokalen (15 Fehler = 10 % der Wortschreibungsfehler, 3 % aller orthografischen Fehler)

- Verwechslung von *ä* und *e*:
zusammän (Kriminalität_15), *benähmen* (Sprache_5), *verschwändet* (Weinen_12), *ver_endert* (Weinen_14), *Lenge* (Weinen_15)
- *i* statt *ie*:
verschidene (Entlohnung_27), *verdinen* (Feminismus_3), *hir* (Kriminalität_12)

Die Belege deuten darauf hin, dass falsche Wortschreibung vor allem dann vorkommt, wenn es eine Diskrepanz zwischen der Schreibweise und der Aussprache oder eine Inkonsequenz der Verwendung von Buchstaben gibt. So ist es z.B. problematisch, wenn die Laut-Buchstaben-Zuordnung nicht eindeutig ist (bei *ä* / *e*) oder ein Buchstabe nicht motiviert zu sein scheint (z.B. *t* in *unterstü_zt*). Homophonie (z.B. *das* vs. *dass*) sowie wortspezifische Verwendungsregeln (z.B. *h* als Dehnungszeichen in *führt*, aber nicht in *spürt*) können beim Rechtschreiberwerb ebenfalls störend sein.

Groß- und Kleinschreibung:

Substantive und Substantivierungen (22 Fehler = 38 % der Groß- und Kleinschreibungsfehler, 5 % aller orthografischen Fehler)

- Substantive:
chancen (Feminismus_13), *gefühl* (Weinen_11)
- substantivierte Adjektive:
nichts neues (Feminismus_3), *im allgemeinen* (Weinen_6), *das wichtigste ist ...* (Wandern_4)
- substantivierte Verben:
mit weinen (Weinen_14)
- Pronomen mit formalen Merkmalen der Substantivierung:
Ich kenne aber sehr Viele, die ... (Feminismus_4), *die Meisten in Peking* (China_10)

Selbstständiger Satz nach Doppelpunkt (16 Fehler = 28 % der Groß- und Kleinschreibungsfehler, 4 % aller orthografischen Fehler)

Ich habe aber viele gute Geschichte gehört: ein Freund von mir hat zum Beispiel ... (Entlohnung_31)

Nationalitätsbezeichnungen¹³ (6 Fehler = 10 % der Groß- und Kleinschreibungsfehler, 1 % aller orthografischen Fehler)

chinesische Großwand (China_1), *Chinesische Rosen* (China_15), *Ungarische Menschen* (Feminismus_3), *englisch lernen* (Sprache_1)

Erwähnenswert sind noch fünf Belege, in denen attributive und prädikative Adjektive großgeschrieben wurden. Die hier aufgezählten Subtypen bzw. Wortarten werden auch in Huber (2023) unter den häufigsten Problemfällen angegeben. Interessant (oder eigentlich unverständlich) ist die falsche Schreibweise von Substantiven, denn die Großschreibung sollte für Lernende auf B2 / C1 schon eine Selbstverständlichkeit sein. Bei anderen substantivisch gebrauchten Wortarten geraten die Studierenden anscheinend oft in Zweifel, ob diese als Substantive aufgefasst werden sollten, weswegen sowohl Klein- als auch Großschreibungsfehler vorkommen. Im Fall der Kleinschreibung nach einem Doppelpunkt lässt sich stark der Einfluss des Ungarischen vermuten.

Getrennt- und Zusammenschreibung:

Substantive (7 Fehler = 41 % der Getrennt- und Zusammenschreibungsfehler, 2 % aller orthografischen Fehler)

IT Firma (Entlohnung_16), *Suffragette Bewegung* (Feminismus_7), *online Seiten* (Feminismus_3), *26 Jährigen* (Weinen_15)

Adverbien (5 Fehler = 29 % der Getrennt- und Zusammenschreibungsfehler, 1 % aller orthografischen Fehler)

- Adverb als Satzglied:
immernoch (Feminismus_4), *wieviel eine Frau ... weint* (Weinen_3)
- Adverb als Verbpartikel:
die Ergebnisse einer neuen Studie dazu rechnet (Weinen_12)

Pronomen (3 Fehler = 18 % der Getrennt- und Zusammenschreibungsfehler, 1 % aller orthografischen Fehler)

die selbe (Kriminalität_8), *soetwas* (Kriminalität_9)

Die meisten Fehler bei der Getrennt- und Zusammenschreibung kommen bei den Substantiven vor. Hier ist es typisch, dass die Glieder eines Kompositums getrennt geschrieben werden. Falsche Zusammenschreibung tritt eher bei Adverbien und Pronomen auf, wenn zwei Wörter wegen ihrer häufigen Kombination als eine feste Fügung betrachtet werden (z.B. *immernoch*, *soetwas*, *wieviel*).

¹³ Obwohl die Schreibweise von Nationalitätsbezeichnungen im Amtlichen Regelwerk (2018) an mehreren Stellen (Substantivierungen allgemein, Sprachen als substantivierte Adjektive, Eigennamen mit nichtsubstantivischen Bestandteilen) gesondert angesprochen wird, schien es mir sinnvoll, all diese Belege in einer Gruppe zusammenzufassen, da die gemeinsame Fehlerquelle die vom Kontext abhängige Groß- und Kleinschreibung von Nationalitätsbezeichnungen ist.

4. Vergleich von Dulko und Falko

In diesem Abschnitt werden die Ähnlichkeiten und Unterschiede zwischen dem bereits analysierten Dulko-Korpus und den 28 Falko-Essaytexten, die zur Analyse ausgewählt wurden (vgl. Abschnitt 2), behandelt. Das Ziel ist festzustellen, wie die orthografische Kompetenz der ungarischen Deutschlernenden in einem Vergleich mit muttersprachlichen Sprechenden positioniert werden kann. Zu diesem Zweck erfolgt eine Gegenüberstellung der relevanten statistischen Daten, während auf eine detaillierte Auflistung von Falko-Belegen verzichtet wird.

Tabelle 3 ermöglicht einen Vergleich der allgemeinen statistischen Daten der orthografischen Fehler in den zwei untersuchten Korpora. Interessanterweise ist die Fehlerhäufigkeit in beiden Fällen 2 %. Die Texte der ungarischen Studierenden sind also orthografisch allgemein auf demselben Niveau wie die Falko-Essays.

	Dulko	Falko
Anzahl der Fehler	450	396
Anteil der Texte mit Fehlern	97 %	100 %
Fehlerhäufigkeit bezogen auf Gesamttokenzahl	2 %	2 %

Tabelle 3
Orthografische Fehler in den untersuchten Korpora insgesamt

Die Verteilung der vier Fehlertypen hinsichtlich der Fehleranzahl zeigt Tabelle 4: Zeichensetzung- und Wortschreibungsfehler kommen im analysierten Falko-Korpus etwas seltener vor, dafür sind Fehler im Bereich der Getrennt- und Zusammenschreibung dreimal so häufig. Das bedeutet, dass die Getrennt- und Zusammenschreibung für die muttersprachlichen Probandinnen und Probanden eine größere Herausforderung darstellt als für die ungarischen Deutschlernenden, wobei die anderen drei Rechtschreibkategorien für beide Sprachbenutzergruppen ca. ähnlich problematisch sind.

	Dulko	Falko
Fehler insgesamt	450 (= 100 %)	396 (= 100 %)
Zeichensetzung	49 %	44 %
Wortschreibung	34 %	30 %
Groß- und Kleinschreibung	13 %	14 %
Getrennt- und Zusammenschreibung	4 %	12 %

Tabelle 4
Verteilung der Fehlertypen in den untersuchten Korpora hinsichtlich Fehleranzahl

Was die Fehler enthaltenden Korpustexte betrifft, so kann man festhalten, dass jeder Fehlertyp im Falko-Korpus in einem höheren Prozentanteil der Texte vorkommt als im Dulko-Korpus (vgl. Tabelle 5). Es gibt keinen fehlerfreien Falko-Essay unter den untersuchten Texten. Der Anteil der Texte mit Fehlern bei der Getrennt- und Zusammenschreibung ist darüber hinaus dreimal so groß im Falko wie im Dulko.

	Dulko	Falko
Texte mit orthografischen Fehlern	62 (= 97 %)	28 (= 100 %)
Texte mit Zeichensetzungsfehlern	89 %	100 %
Texte mit Wortschreibungsfehlern	69 %	82 %
Texte mit Groß- und Kleinschreibungsfehlern	52 %	57 %
Texte mit Getrennt- und Zusammenschreibungsfehlern	22 %	64 %

Tabelle 5
Verteilung der Texte mit den orthografischen Fehlertypen in den untersuchten Korpora

Die Streuungsdaten pro 100 Texttokens (vgl. Tabelle 6) lassen einerseits erkennen, dass es Deutschlernende gab, die besser waren als die besten deutschen Textverfassenden (0 vs. 0,4 Fehler insgesamt; 0 vs. 0,2 Zeichensetzungsfehler). Andererseits gab es ungarische Studierende, die viel schlechter waren als die schlechtesten Muttersprachlerinnen und Muttersprachler (11,8 vs. 7,3 Fehler insgesamt; 7,1 vs. 2,7 Wortschreibungsfehler). Einen etwas größeren Höchstwert der Fehler pro 100 Texttokens erreichten die deutschen Probandinnen und Probanden in den Kategorien Zeichensetzung bzw. Getrennt- und Zusammenschreibung (3,7 und 1,1 Fehler). Durchschnittlich erbrachten die Ungarn aber in jedem Bereich eine ähnliche Leistung wie die Deutschen, was sehr überraschend ist.

	Streuung pro 100 Tokens der Texte		Durchschnitt pro 100 Tokens der untersuchten Korpora	
	Dulko	Falko	Dulko	Falko
Fehler insgesamt	0-11,8	0,4-7,3	2,2	2,2
Zeichensetzung	0-3,3	0,2-3,7	1,1	1,0
Wortschreibung	0-7,1	0-2,7	0,7	0,7
Groß- und Kleinschreibung	0-1,8	0-1,8	0,3	0,3
Getrennt- und Zusammenschreibung	0-0,7	0-1,1	0,1	0,3

Tabelle 6
Streuung und Durchschnitt der orthografischen Fehler in den untersuchten Korpora

Abbildung 3 zeigt die Fehlerhäufigkeit in den untersuchten Falko-Essays pro 100 Tokens der Texte (Durchschnittswerte der Kategorien in Orange) und ermöglicht den Vergleich mit der Streuung im Dulko-Korpus (vgl. Abbildung 2).

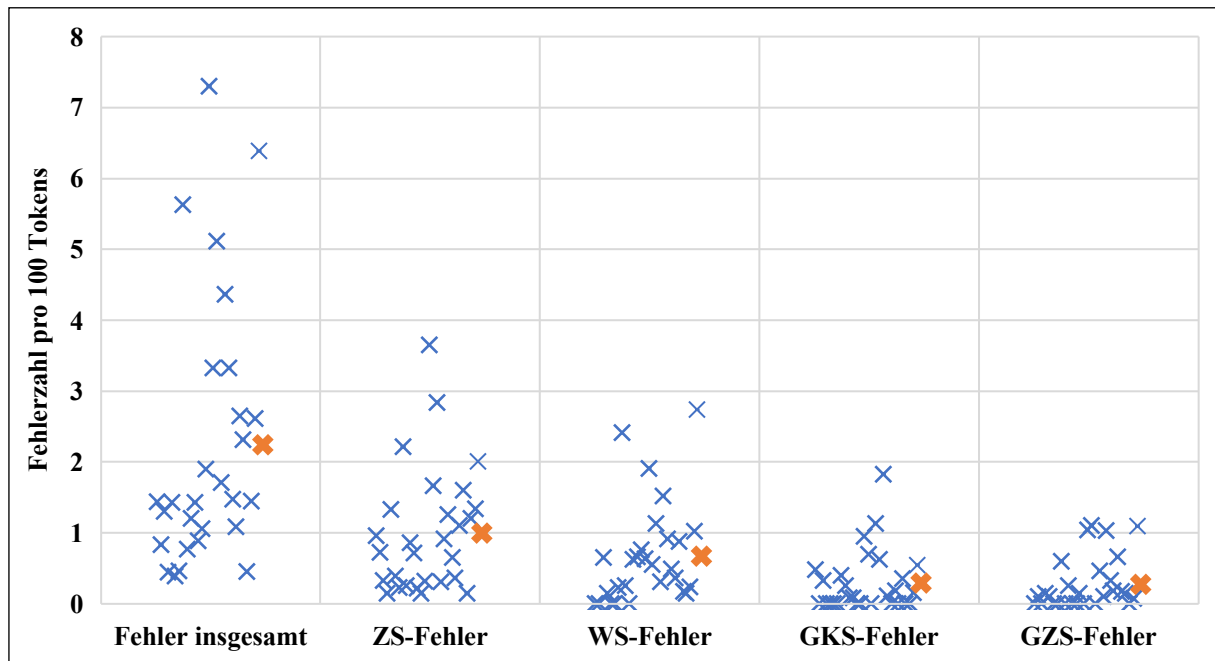


Abbildung 3
Streuung und Durchschnitt der orthografischen Fehler im Falko-Korpus pro 100 Tokens der Texte

Auch hier ist es eindeutig, dass hohe Extremwerte nur in einem kleineren Teil der Texte zu beobachten sind. Die stärkste Streuung weisen die Gesamtfehlerzahlen pro 100 Tokens auf. Diese Streubreite ist aber nicht so erheblich wie im Dulko-Korpus, in dem deutlich mehr Texte über dem Durchschnitt liegen und die zwei höchsten Extremwerte 10 überschreiten, während die Ausreißer im Fall von Falko nicht einmal 8 Fehler pro 100 Tokens erreichen.

Die Spalten der Zeichensetzungsfehler sehen gleich aus (vgl. Abbildung 2 und Abbildung 3): In beiden Korpora gehören 9 % bzw. 7 % der Texte zum oberen Drittel der Werte. Ähnliches gilt für die Fehler bei der Groß- und Kleinschreibung. Bei diesem Fehlertyp sind die meisten Texte im unteren Bereich zwischen 0 und 1, nur ein ganz kleiner Teil der beiden Korpora (3 % bzw. 4 %) befindet sich zwischen 1 und 2. Im Fall der Wortschreibungsfehler gibt es drei Dulko-Texte mit ziemlich hohen Extremwerten (5 % des Korpus), die sich von den restlichen Texten eindeutig absondern. Bei den Falko-Essays enthält das obere Drittel der Streuungsskala (zwischen 1,9 und 2,7) insgesamt 11 % der Essays. Zu demselben Wertebereich gehören nur 6 % der Dulko-Texte. Die Streuung ist jeweils bei der Getrennt- und Zusammenschreibung die kleinste. Den Höchstwert im Dulko (0,7 Fehler pro 100 Tokens) erreichen 2 Texte, während es 4 Falko-Essays mit noch höheren Fehlerquotienten gibt (14 % des untersuchten Falko-Korpus).

Wenn man alle orthografischen Fehler betrachtet, sind die fünf häufigsten Subtypen im Dulko die folgenden:

1. Zeichensetzung: fehlendes Komma (26 % der orthografischen Fehler)
2. Zeichensetzung: überflüssiges Komma (18 % der orthografischen Fehler)
3. Wortschreibung: Fehler im Zusammenhang mit Konsonantenverdopplungen (8 % der orthografischen Fehler)
4. Wortschreibung: Fremdwörter (7 % der orthografischen Fehler)
5. Wortschreibung: Verwendung von einzelnen Konsonanten (5 % der orthografischen Fehler); Groß- und Kleinschreibung: Substantive und Substantivierungen (5 % der orthografischen Fehler)

In den untersuchten Falko-Texten stehen folgende Subtypen auf den ersten fünf Plätzen:

1. Zeichensetzung: fehlendes Komma (34 % der orthografischen Fehler)
2. Wortschreibung: versehentliche Schreibfehler¹⁴ (13 % der orthografischen Fehler)
3. Groß- und Kleinschreibung: Substantive und Substantivierungen (8 % der orthografischen Fehler)
4. Zeichensetzung: überflüssiges Komma (7 % der orthografischen Fehler)
5. Wortschreibung: Fehler im Zusammenhang mit Konsonantenverdopplungen (4 % der orthografischen Fehler)

Abbildung 4 stellt nun die obigen Listen vergleichend dar und zeigt, dass das häufigste Problem sowohl im Dulko als auch im Falko die Nichtsetzung von Kommas ist. Die drei anderen gemeinsamen Subtypen sind in den zwei Korpora in unterschiedlichen Proportionen vertreten. Weitere drei Subtypen sind jeweils korpusspezifisch.

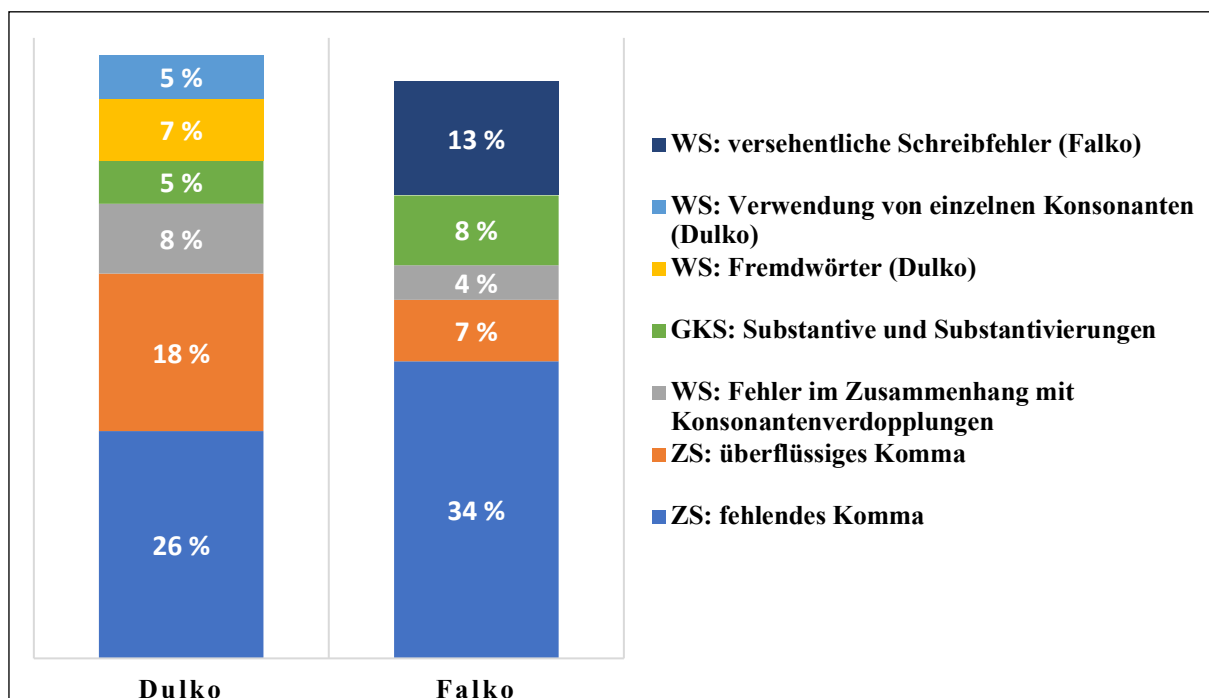


Abbildung 4
Die häufigsten orthografischen Subtypen in den untersuchten Korpora

5. Fazit

Der vorliegende Bericht zielte im Wesentlichen darauf ab, die orthografischen Fehler im Dulko-Korpus statistisch zu erfassen und auszuwerten. Dank der Analyse konnten bedeutende Erkenntnisse über die orthografische Kompetenz von ungarischen Deutschlernenden gewonnen werden, welche sich auch durch den Vergleich mit der muttersprachlichen Kompetenz von Deutschen ergeben.

Zu den wichtigsten Ergebnissen gehört, dass orthografische Fehler in fast allen untersuchten Texten vorkommen, wobei sie nur jeweils 2 % der Gesamttokenzahlen der untersuchten Korpora ausmachen. Die ungarischen Probandinnen und Probanden machten im Durchschnitt nicht mehr Fehler als die deutschen, die Unterschiede auf individueller Ebene sind jedoch groß. Allgemein überwiegen Zeichensetzungs- und Wortschreibungsfehler sowohl im Dulko als auch im Falko, vor allem bei der Kommasetzung, im Zusammenhang mit Konsonantenverdopplungen und der Schreibweise von

¹⁴ Zum Beispiel: *allegmein, gesamtgesellschaftlicher, Mangager, Beispiel.*

Substantiven oder Substantivierungen. Getrennt- und Zusammenschreibungsfehler sind eher in den Falko-Essays relevant. Dadurch erwies sich meine ursprüngliche Hypothese, dass Deutsche generell weniger Fehler machen würden, als falsch.

Der Vergleich mit Huber (2023) ergab, dass ungarische Deutschlernende allgemein die größten Schwierigkeiten mit der Kommasetzung und die wenigsten Probleme mit der Getrennt- und Zusammenschreibung haben. Im Bereich der Groß- und Kleinschreibung sind generell Substantive und Substantivierungen, Nationalitätsbezeichnungen sowie der Satzanfang nach Doppelpunkt besonders problematisch. Andere Analyseergebnisse sind leider entweder korpuspezifisch oder schwer vergleichbar.

Die hier durchgeführte Untersuchung gilt als eine erste allgemeine Auswertung der Fehlerannotation des Dulko-Korpus im Bereich der Orthografie. Auf jeden Fall sind weitere Analysen nötig, um die orthografische Kompetenz der Probandinnen und Probanden des Dulko-Projekts bzw. die Fehlertypen eingehender und präziser beschreiben zu können.

Literatur und Ressourcen

ANNIS. <https://corpus-tools.org/annis/> (01.10.2024).

ANNIS-Suchinterface. <https://korpling.german.hu-berlin.de/annis/> (01.10.2024).

Beeh, Christoph / Drewnowska-Vargáné, Ewa / Kappel, Péter / Modrián-Horváth, Bernadett / Nolda, Andreas / Rauzs, Orsolya / Scheibl, György (2021): *Dulko-Handbuch. Version 1.0*. Szeged: Institut für Germanistik. <https://doi.org/10.14232/dulko-handbuch-v1.0>.

Dulko. <https://arts.u-szeged.hu/kutatas-tudomany/dulko> (01.10.2024).

Falko. <https://hu-berlin.de/falko> (01.10.2024).

Fekete, Olga (2016): *Komplexität und Grammatikalität in der Lernaltersprache. Eine Längsschnittstudie zur Entwicklung von Deutschkenntnissen ungarischer Muttersprachler*. Münster / New York: Waxmann.

Hirschmann, Hagen / Lüdeling, Anke / Shadrova, Anna / Bobeck, Dominique / Klotz, Martin / Akbari, Roodabe / Schneider, Sarah / Wan, Shujun (2022): FALKO. Eine Familie vielseitig annotierter Lernerkorpora des Deutschen als Fremdsprache. In: *Korpora Deutsch als Fremdsprache* 2: 2, 139-148. <https://doi.org/10.48694/kordaf.3552>.

Huber, Ágnes (2023): *Lernaltersprache Deutsch bei ungarischen Germanistik-Studierenden. Eine Untersuchung schriftlicher Sprachprodukte fortgeschrittener Deutschlernender*. Hamburg: Verlag Dr. Kovač.

Modrián-Horváth, Bernadett / Kappel, Péter (2024): Auf der Spur von syntaktischen Fallen in Texten fortgeschrittener Deutschlernender. Zur Verbstellung in Lernendenkorpora. In: *Korpora Deutsch als Fremdsprache* 4: 2.

Molnár, Krisztina (2024): Lernerkorpora in der Tertiärsprachenforschung. In: *Korpora Deutsch als Fremdsprache* 4: 2.

Rat für deutsche Rechtschreibung (2018): *Deutsche Rechtschreibung. Regeln und Wörterverzeichnis*. Mannheim. https://www.rechtschreibrat.com/DOX/rfdr_Regeln_2016_redigiert_2018.pdf (01.10.2024).

Reznicek, Marc / Lüdeling, Anke / Krummes, Cedric / Schwantuschke, Franziska / Walter, Maik / Schmidt, Karin / Hirschmann, Hagen / Andreas, Torsten (2012): *Das Falko-Handbuch. Korpusaufbau und Annotationen Version 2.01*. Berlin: Humboldt-Universität zu Berlin (Institut für deutsche Sprache und Linguistik – Korpuslinguistik). <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/FalkoHandbuchV2/view> (01.10.2024).

Schlotthauer, Susan / Trawiński, Beata (2024): Propositionale Argumente in deutschen Lernertexten. In: *Korpora Deutsch als Fremdsprache* 4: 2.

Szerecz, Sára (2022): Verbvalenzfehler aus dem Übersetzungskorpus von *Dulko*. In: Katona, Tünde (Hrsg.): *Haus als Spiegel - Terézia Moras Aliens - Seiltanz bei Kehlmann - Wüste als Raum - Possessive Attribute im 19. Jh. - Verbvalenzfehler in Übersetzungen. Beiträge Szegeder Germanistikstudierender zur deutschen Literatur- und Sprachwissenschaft sowie zu Deutsch als Fremdsprache*. Szeged: Institut für Germanistik, Universität Szeged, 225-281. <https://doi.org/10.14232/jp.agi.2022.3.6>.

Biographische Notiz: Dr. Orsolya Rauzs ist Oberassistentin am Institut für Germanistik der Universität Szeged (Ungarn). Ihre Promotion hat sie an der Universität Szeged im Bereich Sprachgeschichte abgeschlossen. Sie war wissenschaftliche Mitarbeiterin des Szegeder Dulko-Projekts. Ihre Arbeitsschwerpunkte liegen in den Bereichen Übersetzen und Sprachgeschichte.

Kontaktanschrift:

Dr. Orsolya Rauzs
Universität Szeged
Fakultät für Geistes- und Sozialwissenschaften
Institut für Germanistik
Egyetem u. 2.
6722 Szeged
rauzs@lit.u-szeged.hu

