doi.org/10.36689/uhk/978-80-7435-952-1.005

Sprache und Statistik

Petra LEŽÁKOVÁ
Karlova Univerzita
Pedagogická fakulta
Katedra Germanistiky
Magdalény Rettigové 4, 110 00 Praha 1
petra.lezakova@peuni.cz
ORCID: 0000-0002-5664-7909

Daniel TOTH
Česká zemědělská univerzita v Praze
Institut vzdělávání a poradenství
V Lázních 3, 159 00 Praha 5
tothd@ivp.czu.cz
ORCID: 0000-0002-5993-8170

ABSTRACT

Language and statistics

The article deals with research conducted with prospective German language teachers studying at faculties of education in the Czech Republic. With the help of statistics, hypotheses about the correct proofreading of a German text with mistakes are tested in comparison with native speakers and depending on when the students started learning German and whether they learnt German before English.

KEYWORDS

German language, statistics, faculties of education in the Czech Republic, German native speakers, correction of mistakes

1. Einführung

Sprache und Statistik – zwei verschiedene Welten, die scheinbar nichts miteinander zu tun haben, sich aber gegenseitig unterstützen und ergänzen. Metaphorisch kann man sagen, dass Statistik wie ein Zweig aus dem Baum der Linguistik wächst und unser sprachliches Wissen erweitert.

Mithilfe von aufgestellten Hypothesen und deren Verifizierung oder Falsifizierung können wir auch in der Linguistik quantitativ forschen und dank der Statistik wertvolle Erkenntnisse gewinnen.

Das Ziel des vorliegenden Artikels ist die Beantwortung von fünf Forschungsfragen auf der Grundlage von Hypothesen und deren statistischer Überprüfung, die wir im Rahmen einer Befragung von Hochschulstudierenden, die als künftige Deutschlehrkräfte an pädagogischen Einrichtungen in der Tschechischen Republik arbeiten möchten, stellten. Die Ergebnisse werden von Deutschstudierenden in der Tschechischen Republik mit den Ergebnissen von MuttersprachlerInnen in Deutschland und Österreich verglichen.

In den folgenden Abschnitten werden ähnliche Fragen erörtert wie die unsrigen, die zuvor von anderen AutorInnen untersucht wurden und deren Ergebnisse mit den unsrigen diskutiert werden.

Kühn (1988:44–49) stellte 276 ausländischen Deutschlehrkräften der Universität Leipzig einen Text mit 20 Fehlern zum Korrekturlesen zur Verfügung. Für jede richtige Korrektur wurden +2 Punkte vergeben, für jede falsche Korrektur –2 Punkte, für jede richtig ersetzte korrekte Form –1 Punkt. Insgesamt konnten 40 Punkte erreicht werden, die keine der Lehrkräfte erreichte; das beste Ergebnis betrug 37 Punkte. Nur 3 % der Lehrkräfte erzielten 30 oder mehr Punkte, 28 % erreichten weniger als 10 Punkte, 5 % der Lehrkräfte unterschritten sogar die Negativschwelle. Die Forschung zeigt, dass selbst der Lehrberuf keine Garantie für einwandfreie Sprachbeherrschung ist.

Mit der Fehlerkorrektur befasste sich auch Ondráková (2014:122–128). Ihre Forschungsstichprobe bestand aus 140 Personen – 55 Deutschlehrkräften und 85 Deutschstudierenden – denen sie sieben deutsche Sätze mit 18 Fehlern zum Korrekturlesen gab. Von den 2 520 Fehlern wurde weniger als die Hälfte richtig korrigiert; 47,54 % Fehler. Ondráková kommt zu dem Schluss, dass Unterrichtserfahrung keine ausreichende Qualifikation für die Korrektur von Fehlern ist.

An Kalkans Studie (2021:103–110) nahmen 37 angehende Deutschlehrkräfte, Studierende im sechsten Semester, teil, die sich ein Semester lang mit Fehleranalyse beschäftigt hatten. Die Studierenden wurden in zwei Gruppen eingeteilt; 30 und sieben Studierende. Die kleinere Gruppe bestand aus Studierenden, die aus Deutschland zurückgekehrt waren, d. h., sie begegneten der Fremdsprache in ihrer natürlichen Umgebung und hatten daher bessere Sprachkenntnisse. Diese Gruppe wurde als Gruppe 2 bezeichnet. Die größere Gruppe, Gruppe 1, bestand aus Studierenden, die einen einjährigen studienvorbereitenden Deutschkurs an der Hochschule absolviert hatten.

Der Stichprobe von Studierenden wurden zwei Texte mit Fehlern vorgelegt. Im Text 1 mussten sie die Fehler identifizieren und sie dann anhand des vorgelegten Rasters klassifizieren. Im Text 2 waren die Fehler bereits markiert und nur noch deren Klassifizierung erforderlich.

Text 1 umfasste 144 Wörter und wurde von einem Studierenden im zweiten Semester verfasst; Kalkan modifizierte den Text teilweise. Die Studierenden beider Gruppen mussten die Fehler des Textes anhand der beigefügten Vorlage identifizieren. Fehler konnten in die Kategorie Rechtschreib-, Wortwahl-, Kasus-, Numerus-, Geschlechts-, Übereinstimmungs- und Deklinationsfehler fallen. Der Text enthielt 35 Fehler. Nach 40 Minuten wurde Text 1 gesammelt und Text 2 den Studierenden präsentiert, in dem Fehler bereits markiert wurden; die Studierenden sollten sie kategorisieren.

Text 1 enthielt 35 Fehler, es handelte sich um insgesamt 1295 Fehler (35 x 37), falls alle Fehler von allen Studierenden erkannt worden wären. Im Text 1 wurden nur 66 % der Fehler festgestellt. Von zu entdeckten Fehlern identifizierte Gruppe 2 86 % der Fehler, Gruppe 1 identifizierte 61 % der Fehler. Die höchste Erfolgsquote bei der Fehlererkennung lag für beide Gruppen bei orthografischen Fehlern bei 70 %. Die am zweithäufigsten identifizierte Fehlergruppe waren Kongruenzfehler (63 %). Die dritte Gruppe bestand aus Deklinationsfehlern, 59 %. Lexikalische Fehler wurden von den Studierenden bei 46 % entdeckt, gefolgt von Fehlern in der Groß- und Kleinschreibung (45 %) und Fehlern im Numerus (42 %). Die wenigsten Fehler stellten Studierende bei Geschlechtsfehlern fest, nämlich nur 20 %.

Gruppe 2 identifizierte die meisten Fehler. Der größte Unterschied zwischen den Gruppen bestand bei Fehlern im Geschlecht und bei lexikalischen Fehlern. Die Gruppen lagen bei der Bestimmung von Fehlern in Kasus und Deklination am nächsten beieinander. Am Ende der Arbeit mit Text 1 zeichnete sich eine neue Erkenntnis ab, nämlich dass die Studierenden bei der Korrektur von Fehlern auch Fehler identifizierten, die keine Fehler waren. Solche „Fehler“ wurden in 187 Fällen festgestellt, was einem Durchschnitt von fünf Fehlern pro Studierenden entspricht. Bei Gruppe 1 lag diese Zahl bei 5,5, bei Gruppe 2 bei 3,3.

Die Arbeit mit Text 2 bestand darin, zu überprüfen, ob Sprachkenntnisse mit der Fähigkeit, die Art von Fehlern zu bestimmen, korrelieren. Überraschenderweise wurde festgestellt, dass der Unterschied in der Fehlerklassifizierung zwischen den beiden Gruppen nur 4 % Prozent betrug; bei der Fehlererkennung waren es 25 %. Somit spielt das Sprachniveau der Studierenden eine größere Rolle bei der Erkennung von Fehlern als bei deren Einordnung.

2. Methodik

2.1 Forschungsstichprobe

Unsere Forschung an pädagogischen Fakultäten in der Tschechischen Republik fand in den Sommersemestern der Studienjahre 2021/2022 und 2022/2023 statt, jeweils im sechsten Semester des Bachelorstudiums. Wir wählten das letzte Semester des Bachelorstudiums bewusst, weil wir davon ausgingen, dass die Studierenden am Ende ihres Bachelorstudiums bereits mit allen Phänomenen vertraut sind, denen sie als zukünftige Deutschlehrkräfte begegnen könnten.

Die Teilnehmenden der Untersuchung waren Studierende aus sechs pädagogischen Fakultäten, an denen die deutsche Sprache gelehrt wird: nämlich der Karls-Universität Prag, der Universität Hradec Králové, der Palacký-Universität Olomouc, der Südböhmischen Universität České Budějovice, der Westböhmischen Universität Plzeň und der Technischen Universität Liberec.

Insgesamt nahmen 194 Studierende im Laufe der Jahre an unserer Forschung teil. Sie bekamen einen deutschen Text mit acht Sätzen auf B2-Niveau, der 30 Fehler beinhaltete. Die Aufgabe der Studierenden war es, diese Fehler zu finden und zu korrigieren. Dabei handelte es sich um morphologische, syntaktische, lexikalische und orthografische Fehler. Für die Korrektur der Fehler gab es kein Zeitlimit. Die meisten Studierenden schafften es, den Text innerhalb von 15 Minuten zu korrigieren.

Zusammen mit dem deutschen Text wurde den Studierenden ein Fragebogen ausgehändigt, in dem sie zu weiteren Fragen im Zusammenhang mit unserer Forschung befragt wurden: U. a. wann sie mit dem Deutschlernen begannen und ob Deutsch ihre erste oder weitere Fremdsprache sei.

Das Forschungsteam ging davon aus, dass das Abschlussniveau des Bachelorstudiums für Studierende aller teilnehmenden Universitäten ähnlich sei und da wir keine Unterschiede zwischen einzelnen Universitäten machen wollten, werteten wir die erzielten Ergebnisse als Ganzes aus.

In den gleichen Jahren, in denen unsere Forschung an den pädagogischen Fakultäten in der Tschechischen Republik durchgeführt wurde, wurden auch Forschungen an der Katholischen Universität Eichstätt-Ingolstadt und an der Pädagogischen Hochschule Steiermark durchgeführt. Den ortsansässigen Studierenden wurde der gleiche deutsche Text mit Fehlern zugeteilt wie den tschechischen Studierenden. Über einen Zeitraum von zwei Jahren holten wir Textkorrekturen von insgesamt 81 Studierenden aus Deutschland und Österreich (überwiegend aus dem zweiten Semester) ein, wobei wir auch nicht zwischen den Studierenden differenzierten und diese in unserem Artikel als MuttersprachlerInnen bezeichnen.

2.2 Verwendete statistische Tests

Im weiteren Text beschreiben wir die Gründe und die Anwendung der in der Forschung verwendeten Methoden, einschließlich statistischer Analysen und beschreibender Merkmale.

Zuerst mussten wir den Shapiro-Wilk-Test durchführen, um festzustellen, ob die Daten normalverteilt sind und die Bedingung der Normalität entsprechen (Baayen, 2008:11). Dieser

Test ist für kleine bis mittelgroße Datensätze erforderlich und gilt als einer der leistungsfähigsten Normalitätstests. Der Shapiro-Wilk-Test berechnet den W-Wert und den p-Wert, die mit dem Signifikanzniveau α, normalerweise 0,05, verglichen werden. Die Nullhypothese des Tests behauptet, dass die Daten normalverteilt seien. Wenn der p-Wert kleiner als 0,05 ist, lehnen wir die Nullhypothese ab und schließen daraus, dass die Daten nicht normalverteilt sind. Umgekehrt, wenn der p-Wert größer oder gleich 0,05 ist, lehnen wir die Nullhypothese nicht ab und betrachten die Daten als normalverteilt. Wenn die Daten die Normalbedingung nicht erfüllen, fahren wir mit dem Testen mit einem nichtparametrischen Mann-Whitney-U-Test fort, der kein normales Datenlayout erfordert (Zhang, 2023).

Auf diese Weise stellten wir sicher, dass wir je nach Verteilung unserer Daten einen geeigneten statistischen Test verwenden. In unserer Forschung wird der Mann-Whitney-U-Test angewendet, bei dem es sich um einen nichtparametrischen Test handelt, der auch als Mann-Whitney-Wilcoxon-Test oder Wilcoxon-Test bekannt ist. Dieser Test wird verwendet, um die Unterschiede zwischen zwei unabhängigen Gruppen zu vergleichen, und er ist besonders nützlich, wenn die Daten nicht normalverteilt sind und unterschiedliche Stichprobengrößen aufweisen. Zu Beginn des Tests formulierten wir die Null- und Alternativhypothese:

Nullhypothese (H_0): Die Mediane der beiden Gruppen sind gleich.

Alternativhypothese (H_A): Die Mediane der beiden Gruppen sind unterschiedlich.

Das Testverfahren umfasste mehrere Schritte. Zuerst sammelten wir Daten von zwei unabhängigen Stichproben von zu analysierenden Daten. Anschließend kombinierten wir beide Proben zu einem Satz und ordneten alle Beobachtungen in aufsteigender Reihenfolge an. Wir wiesen jeder Beobachtung eine Sequenz zu. Sind die Werte gleich (sogenannte Links), wiesen wir ihnen eine durchschnittliche Reihenfolge zu. Wir berechneten die Summe des Rangs für jede der Stichproben. Anschließend führten wir die Berechnung der U-Statistik durch. Wir setzten die Rangsummen für beide Stichproben in die Formel für U ein. Dadurch erhielten wir Werte, mit denen wir feststellten, ob es einen statistisch signifikanten Unterschied zwischen zwei unabhängigen Gruppen gibt. Dieses Verfahren ermöglichte eine gründliche Bewertung der Unterschiede zwischen den beobachteten Gruppen und half, das untersuchte Phänomen genauer zu verstehen.

$$U1 = n1 \times n2 + \frac{n1(n1+1)}{2} - R1$$

$$U2 = n1 \times n2 + \frac{n2(n2+1)}{2} - R2$$

Wenn wir es in die Formel einsetzten, erhielten wir die Werte von U1 und U2, wobei n1 und n2 die Stichprobengrößen und R1 und R2 die Reihensummen der Stichproben bezeichnen. Wir wählten die kleinere von U1 und U2 als Teststatistik aus. Der kritische Wert von U kann aus den Tabellen der kritischen Werte für ein bestimmtes Signifikanzniveau und einen bestimmten Stichprobenumfang ermittelt werden. Wenn der berechnete U-Wert kleiner oder gleich dem kritischen Wert ist, lehnen wir die Nullhypothese ab. Dann verwendeten wir die Z-Statistik, die wir in den Ergebnissen interpretieren werden. Die Formel zur Berechnung der Z-Statistik im Mann-Whitney-U-Test lautet:

$$Z = \frac{U - \mu_U}{\sigma_U}$$

Dabei ist U der berechnete Wert der Statistik U, der Zähler ist der berechnete Wert der Statistik U abzüglich des erwarteten Wertes der Statistik. Im Nenner steht die Standardabweichung der Statistik U. Wenn der Stichprobenumfang groß genug ist (in der Regel ist eine Stichprobengröße n größer als 20 erforderlich), kann die Statistik U durch eine Normalverteilung angenähert werden, was die Berechnung der Z-Statistik und anschließend des p-Wertes aus der Normalverteilung ermöglicht. Die Interpretation der Z-Statistik im Mann-Whitney-U-Test liefert wichtige Informationen über die Unterschiede zwischen zwei Gruppen. Ein positiver Z-Wert zeigt an, dass der Mittelwert der Rangfolge in Gruppe 1 höher ist als in Gruppe 2, was darauf hindeutet, dass die erstgenannte Gruppe eine höhere Rangfolge aufweist. Umgekehrt deutet ein negativer Z-Wert auf die gegenteilige Tendenz hin, d. h. der durchschnittliche Rang in Gruppe 2 ist höher als in Gruppe 1, was auf das Vorhandensein höherer Werte in der letzten Gruppe hindeutet. Die Größe der Z-Statistik gibt dann Auskunft darüber, wie weit die beobachteten Unterschiede von den erwarteten Werten unter der Bedingung der Gültigkeit der Nullhypothese entfernt sind. Je größer der absolute Wert von Z ist, desto mehr Beweise liegen gegen die Nullhypothese vor.

Die Signifikanz der Z-Statistik wird außerdem im Hinblick auf das gewählte Signifikanzniveau bewertet; wenn der absolute Wert von Z größer ist als der kritische Wert für das gegebene Signifikanzniveau, verwerfen wir die Nullhypothese zugunsten der Alternativhypothese. Der p-Wert, sofern vorhanden, liefert uns ebenfalls einen wichtigen Indikator; ein niedriger p-Wert deutet auf eine größere Wahrscheinlichkeit hin, dass die beobachteten Unterschiede statistisch signifikant sind. Auf diese Weise interpretiert, ermöglichen die Ergebnisse der Z-Statistik eine objektive Bewertung des Vorhandenseins und der Signifikanz von Unterschieden zwischen den beiden verglichenen Gruppen.

Der p-Wert ist ein Schlüsselelement für die Interpretation der Ergebnisse des Mann-Whitney-U-Tests. Er gibt die Wahrscheinlichkeit an, dass die beobachteten Unterschiede zwischen den beiden Gruppen zufällig auftreten könnten, wenn die Nullhypothese wahr wäre. Ist der p-Wert größer als ein vorgegebenes Signifikanzniveau, liegen keine ausreichenden Beweise gegen die Nullhypothese (H_0) vor. In diesem Fall lehnen wir die Nullhypothese nicht ab und schließen daraus, dass die Unterschiede zwischen den Medianen der beiden Gruppen statistisch nicht signifikant sind. Liegt der p-Wert unter dem Signifikanzniveau, wird die Nullhypothese verworfen. D. h., die Unterschiede zwischen den Medianen der beiden Gruppen sind statistisch signifikant, was die Alternativhypothese (H_A) unterstützt. Genauer gesagt:

p-Wert > 0,05: Wir lehnen H_0 nicht ab (die Mediane der beiden Gruppen sind gleich).

p-Wert ≤ 0,05: Wir lehnen H_0 ab und akzeptieren H_A (die Mediane der beiden Gruppen sind unterschiedlich).

Somit hilft der p-Wert bei der Entscheidung, ob die beobachteten Unterschiede zwischen den Gruppen stark genug sind, um als statistisch signifikant angesehen zu werden, oder ob sie auf zufällige Variationen in den Daten zurückzuführen sind.

In unserer Studie verwendeten wir auch den Korrelationskoeffizienten nach Spearman, um die Beziehungen zwischen den uns interessierenden Variablen zu bewerten. Die Primärdaten wurden mithilfe des Shapiro-Wilk-Tests auf ihre Normalverteilung geprüft. Die Ergebnisse zeigten, dass die Daten nicht normalverteilt waren, was zur Wahl vom Spearman-Korrelationskoeffizienten führte, der ebenfalls keine Normalverteilung voraussetzt.

Der Korrelationskoeffizient drückt die Stärke und Richtung der Beziehung zwischen zwei Variablen aus. Ein Wert nahe bei 1 weist auf eine starke positive Korrelation hin, bei der sich die beiden Variablen in dieselbe Richtung bewegen. Ein Wert nahe bei –1 zeigt eine starke negative Korrelation an, bei der sich die Variablen in entgegengesetzte Richtungen bewegen. Ein Wert nahe 0 bedeutet, dass es keine Korrelation zwischen den Variablen gibt. Die folgenden Hypothesen werden zur Prüfung der Korrelation verwendet:

Nullhypothese (H_0): Es gibt keine Korrelation zwischen zwei Variablen.

Alternativhypothese (H_A): Es besteht eine Korrelation zwischen zwei Variablen.

Der Spearman-Korrelationskoeffizient wird verwendet, um eine Beziehung zwischen zwei Variablen zu messen, was bedeutet, dass die Variablen jede Art von Korrelation aufweisen können. Die Interpretation der Ergebnisse des Korrelationstests bestimmt, ob die Nullhypothese verworfen werden kann und die Alternativhypothese, die besagt, dass es eine signifikante Korrelation zwischen den Variablen gibt, akzeptiert werden kann.

3. Ergebnisse

1. Forschungsfrage

Alle Berechnungen unserer Forschung wurden mit dem Online-Rechner Statistics Kingdom (www.statskingdom.com) durchgeführt. Zuerst handelte es sich um einen Hypothesentest auf der Grundlage der ersten Forschungsfrage, in der wir die Ergebnisse des Korrekturlesens durch MuttersprachlerInnen und Studierende an tschechischen Universitäten vergleichen wollten:

> *Gibt es einen statistisch signifikanten Unterschied zwischen der Anzahl der richtigen Korrekturen von Studierenden an tschechischen Universitäten und denen aus Deutschland und Österreich?*

Zunächst wurde ein Shapiro-Wilk-Test durchgeführt, um festzustellen, ob die Daten normalverteilt sind. Es wurde ein Signifikanzniveau α von 0,05 verwendet. Die Forschung umfasste 194 Studierende aus Tschechien und 81 Studierende aus Deutschland und Österreich, um eine ausreichende Anzahl von Beobachtungen zu gewährleisten. Zuerst testeten wir die Gruppe von MuttersprachlerInnen auf Normalität und stellten die folgenden Null- und Alternativhypothesen auf:

H_0: Die Daten sind normalverteilt.

H_A: Die Daten sind nicht normalverteilt.

Die Nullhypothese (H_0) wurde verworfen, weil der p-Wert kleiner als α war, was bedeutet, dass die Daten nicht normalverteilt sind. D. h., dass der Unterschied zwischen den Stichprobendaten und der Normalverteilung statistisch signifikant ist. Der p-Wert wurde mit 0,004654 berechnet. Je kleiner der p-Wert ist, desto mehr spricht er für die Alternativhypothese.

Die W-Statistik erreichte einen Wert von 0,9528, was nicht innerhalb des Annahmebereichs von 95 % liegt, der durch Werte von 0,9695 bis 1 definiert ist. Zusammenfassend lässt sich sagen, dass eine der beiden Forschungsgruppen keine Normalverteilung aufwies. Daher wurde der Mann-Whitney-Test, der keine Normalverteilung voraussetzt, zur Analyse verwendet. Es handelt sich um einen nichtparametrischen Test, der für unsere Zielsetzung geeignet ist. Die Gruppen haben nicht notwendigerweise die gleiche Anzahl von Beobachtungen, der U-Test vergleicht die Zufallsvariablen der beiden Gruppen. Das Signifikanzniveau α wurde auf 0,05 festgelegt.

Da der berechnete p-Wert $0 < \alpha$ ist, bedeutet dies, dass H_0 verworfen wird. Die statistische Interpretation sagt, dass die ausgewählten Werte der tschechischen Studierenden und der MuttersprachlerInnen-Stichprobe als statistisch signifikant unterschiedlich angesehen werden können. Der Unterschied zwischen dem zufällig ausgewählten Wert der Stichprobe der tschechischen Studierenden und der MuttersprachlerInnen ist groß genug, um statistisch signifikant zu sein. Der p-Wert bedeutet, dass die Wahrscheinlichkeit eines Fehlers vom Typ I (Zurückweisung einer korrekten H_0) gering ist: je kleiner der p-Wert, desto mehr bestätigt er die H_A.

Die Z-Test-Statistik beträgt –12,0118; der Wert ist nicht im Annahmebereich von 95 %. Der Wert der U-Statistik beträgt 649,5, was ebenfalls außerhalb des Annahmebereichs von 95 % liegt, der durch Werte von 6681,0349 bis 9032,9651 gekennzeichnet ist. Dies bedeutet

wiederum, dass ein statistisch signifikanter Unterschied in dem statistisch analysierten Merkmal besteht.

Die erste Forschungsfrage wurde mit Ja beantwortet: Es besteht ein statistisch signifikanter Unterschied zwischen der Anzahl richtiger Korrekturen von Studierenden an tschechischen Universitäten und Studierenden aus Deutschland und Österreich.

Studierende aus Tschechien korrigierten fast ein Viertel aller Fehler richtig, Studierende aus Deutschland und Österreich entdeckten zwei Drittel aller Fehler.

2. Forschungsfrage

Zur Beantwortung der zweiten und dritten Forschungsfrage unterteilten wir die Fehler im deutschen Text in lexikalische und orthografische Fehler (die wir zu einer Gruppe zusammenfassten) sowie in morphosyntaktische Fehler. Das Forschungsteam interessierte, wie die Fehlerkorrektur von MuttersprachlerInnen im Vergleich zu den von Studierenden an tschechischen Universitäten ausfiel. Die zweite Forschungsfrage formulierten wir wie folgt:

> *Gibt es einen statistisch signifikanten Unterschied zwischen der Anzahl der richtigen Korrekturen von lexikalischen und orthografischen Fehlern von Studierenden an tschechischen Universitäten und denen aus Deutschland und Österreich?*

Der zur Korrektur eingereichte deutsche Text enthielt 14 lexikalische und orthografische Fehler. Aufgrund der geringen Anzahl der Messungen wurde das Signifikanzniveau α im durchgeführten Mann-Whitney-U-Test auf 0,1 festgelegt.

H_0: Es gibt keinen statistisch signifikanten Unterschied zwischen den beiden Gruppen.

H_A: Es gibt einen statistisch signifikanten Unterschied zwischen den beiden Gruppen.

Da der p-Wert > α ist, kann H_0 nicht verworfen werden. Der zufällig ausgewählte Wert der Gruppe der tschechischen Studierenden ist gleich dem zufällig ausgewählten Wert der Gruppe der MuttersprachlerInnen. Der Unterschied zwischen dem zufällig ausgewählten Wert der tschechischen und der muttersprachlichen Studierenden ist nicht groß genug, um statistisch signifikant zu sein.

Der p-Wert beträgt 0,6458, was bedeutet, dass die Wahrscheinlichkeit eines Fehlers vom Typ I, d. h. der Ablehnung einer korrekten H_0, zu hoch ist: 0,6458 (64,58 %). Der p-Wert ist umso größer, je mehr er für H_0 spricht.

Die Z-Test-Statistik steht bei –0,4597, was innerhalb des Annahmebereichs von 90 % liegt. Die U-Statistik beträgt 87,5, was sich auch innerhalb des Annahmebereichs von 90 % befindet, der Werte von 62,2163 bis 133,7837 beinhaltet.

Die zweite Forschungsfrage wurde verneint. Bei der Korrektur von lexikalischen und orthografischen Fehlern ist es statistisch nicht aussagekräftig, ob sie von tschechischen Studierenden oder Studierenden aus Deutschland oder Österreich durchgeführt wird.

Dieses Ergebnis überraschte uns, weil wir erwarteten, dass MuttersprachlerInnen lexikalische und orthografische Fehler deutlich besser korrigieren werden als Studierende tschechischer Universitäten.

3. Forschungsfrage

Die dritte Forschungsfrage lautete wie folgt:

> *Gibt es einen statistisch signifikanten Unterschied zwischen der Anzahl der richtigen Korrekturen von morphosyntaktischen Fehlern von Studierenden an tschechischen Universitäten und denen aus Deutschland und Österreich?*

Der zur Korrektur eingereichte deutsche Text enthielt 16 morphosyntaktische Fehler. Aufgrund der geringen Anzahl der Messungen wurde das Signifikanzniveau α im durchgeführten Mann-Whitney-U-Test auf 0,1 festgelegt.

H_0: Es gibt keinen statistisch signifikanten Unterschied zwischen den beiden Gruppen.

H_A: Es gibt einen statistisch signifikanten Unterschied zwischen den beiden Gruppen.

Da der berechnete p-Wert kleiner als α ist, wird die Nullhypothese abgelehnt, d. h. der zufällig ausgewählte Wert der Gruppe der tschechischen Studierenden ist nicht gleich dem zufällig ausgewählten Wert der Gruppe der MuttersprachlerInnen. Der Unterschied zwischen dem zufällig ausgewählten Wert der Gruppe der tschechischen Studierenden und der Gruppe der MuttersprachlerInnen ist groß genug, um statistisch signifikant zu sein.

Die Nullhypothese wird abgelehnt und die Alternativhypothese wird bei einem Signifikanzniveau von 10 % angenommen. Der p-Wert beträgt 0,04561, was bedeutet, dass die Wahrscheinlichkeit eines Fehlers vom Typ I (Ablehnung einer korrekten H_0) gering ist: 0,04561 (4,56 %). Je kleiner der p-Wert, desto eher spricht er für H_A.

Die Z-Test-Statistik beträgt –1,999 und liegt damit nicht im Annahmebereich von 90 %. Auch die U-Statistik (U=90) 74,5 befindet sich nicht im Annahmebereich von 90 %, der sich von 84,3891 bis 171,6109 erstreckt.

Im Gegensatz zur zweiten Forschungsfrage wurde die dritte Forschungsfrage bejaht. Es wurde festgestellt, dass bei der Korrektur morphosyntaktischer Fehler ein statistisch signifikanter Unterschied zwischen den Korrekturen von Studierenden an tschechischen Universitäten und denen von Studierenden aus Deutschland und Österreich besteht.

Dieses Ergebnis entsprach unseren Erwartungen; wir gingen davon aus, dass MuttersprachlerInnen besser in der Lage sein werden, Fehler in der Grammatik und der Satzstruktur zu erkennen als Studierende tschechischer Universitäten.

4. Forschungsfrage

In der vierten Forschungsfrage wollten wir herausfinden, ob es für die Fehlerkorrektur eine Rolle spielt, inwieweit Studierende an tschechischen Universitäten früher mit dem Deutsch- oder Englischstudium begannen. Die Forschungsfrage lautete wie folgt:

> *Gibt es einen statistisch signifikanten Unterschied zwischen der Anzahl der richtigen Korrekturen von Studierenden an tschechischen Universitäten, die Deutsch vor Englisch lernten, und denen, die Englisch vor Deutsch lernten?*

Es wurde ein Zwei-Stichproben-Proportionstest mit zwei Stichproben verwendet, das Signifikanzniveau α betrug 0,05 (5 %). Die Forschungsstichprobe der tschechischen Studierenden wurde in zwei Gruppen aufgeteilt, der Durchschnitt der Anzahl der richtigen Antworten für jede Gruppe wurde mit dem besten Ergebnis der Gruppe verglichen.

H_0: Es gibt keinen Unterschied zwischen den beiden Gruppen.

H_A: Es gibt einen Unterschied zwischen den beiden Gruppen.

Da der p-Wert größer als α berechnet wurde, kann H_0 nicht verworfen werden. Es wird angenommen, dass der Anteil der Gruppe, die mit Deutsch früher als mit Englisch anfing, gleich der Proportion der Gruppe, deren frühere Fremdsprache Englisch war, ist. Der Stichprobenunterschied zwischen den Anteilen beider Gruppen ist nicht groß genug, um statistisch signifikant zu sein. Die Nullhypothese kann nicht verworfen werden.

Der p-Wert beträgt 0,324663, was bedeutet, dass die Wahrscheinlichkeit eines Fehlers vom Typ I, d. h. der Ablehnung der richtigen H_0, zu hoch ist: 0,3247 (32,47 %). Je größer der p-Wert, desto eher spricht er für H_0.

Die Z-Statistik ist gleich 0,984922 und liegt damit innerhalb des Annahmebereichs von 95 %, der sich zwischen –1,959964 und 1,959964 bewegt. Das Ergebnis der Verhältnisse p1 und p2 ist 0,14 und befindet sich somit im Annahmebereich von –0,282576 bis 0,282576.

Die vierte Forschungsfrage wurde verneint. Für die Anzahl richtiger Korrekturen ist es statistisch nicht signifikant, ob sie von Studierenden vorgenommen wurden, deren erste Fremdsprache Deutsch oder Englisch war.

5. Forschungsfrage

In der fünften Forschungsfrage wollten wir das Alter berücksichtigen, in dem die tschechischen Studierenden mit dem Erlernen der deutschen Sprache begannen; wir formulierten die Frage wie folgt:

> *Hängt die Anzahl der richtigen Korrekturen von Studierenden an tschechischen Universitäten vom Alter ab, in dem sie mit dem Deutschlernen begannen?*

Es wurde eine Korrelationsanalyse durchgeführt, wobei die Bedingungen der Normalverteilung nicht erfüllt waren, was durch den Shapiro-Wilk-Test überprüft wurde. Das Signifikanzniveau α wurde auf 0,05 festgelegt.

H_0: Es besteht keine Korrelation zwischen der Anzahl der richtigen Korrekturen und dem Alter, in dem die Studierenden mit dem Deutschlernen begannen.

H_A: Es besteht eine statistisch signifikante Korrelation zwischen der Anzahl der richtigen Korrekturen und dem Alter, in dem die Studierenden mit dem Deutschlernen begannen.

Da der p-Wert kleiner als α ist, wird H_0 abgelehnt. Die Annahme ist, dass die Stichprobenkorrelation nicht gleich der erwarteten Korrelation ist; mit anderen Worten, dass der Unterschied zwischen der Stichprobenkorrelation und der erwarteten Korrelation groß genug ist, um statistisch signifikant zu sein.

Der p-Wert beträgt 0,006897, was bedeutet, dass die Wahrscheinlichkeit eines Fehlers vom Typ I (Ablehnung der richtigen H_0) gering ist: 0,006897 (0,69 %). Je kleiner der p-Wert ist, desto eher spricht er für H_A.

Die Teststatistik T ist gleich –2,7313, was nicht innerhalb des Annahmebereichs von 95 % liegt, der sich von –1,9724 bis 1,9724 erstreckt.

Der Spearman-Korrelationskoeffizient beträgt jedoch –0,1934: Es besteht eine Korrelation/Abhängigkeit, die sehr schwach und statistisch nicht signifikant ist.

Auch die fünfte Forschungsfrage wurde verneint. Es konnte keine statistisch signifikante Abhängigkeit richtiger Korrekturen vom Alter, in dem Studierende tschechischer Universitäten mit dem Deutschstudium begannen, nachgewiesen werden.

4. Diskussion

Im Bereich der analytischen Linguistik erwiesen sich quantitative Methoden als wesentlich für die Analyse linguistischer Daten (Gries, 2009). Auch Baayen (2008) bietet eine praktische Einführung in die Statistik mit R, die für die Analyse linguistischer Daten nützlich ist. Baroni und Evert (2009) untersuchen statistische Methoden für die Verwendung von Textkorpora, die ein tieferes Verständnis und eine bessere Nutzung von linguistischen Daten ermöglichen. Biber und Jones (2009) heben die Bedeutung quantitativer Methoden in der analytischen Linguistik hervor und bieten detaillierte Ansätze für deren Anwendung.

Aus den Forschungen von Kühn und Ondráková geht hervor, dass das Korrigieren von Fehlern keine einfache Angelegenheit ist und auch Deutschlehrkräfte dabei Fehler machen.

Unsere Forschung zeigte, dass die Fehlerkorrektur selbst für MuttersprachlerInnen nicht immer einfach ist, insbesondere wenn es um die Korrektur lexikalischer und orthografischer Fehler geht. Eine Erklärung könnte sein, dass selbst MuttersprachlerInnen einen Rechtschreibfehler leichter übersehen können als einen grammatikalischen oder syntaktischen Fehler.

Unsere Forschung ergab keinen signifikanten Unterschied in den Korrekturen lexikalischer und orthografischer Fehler zwischen Studierenden von tschechischen Universitäten und MuttersprachlerInnen.

Im Gegensatz hierzu steht Kalkans Forschung, die einen Unterschied in den Korrekturen in der Gruppe der lexikalischen Fehler zwischen Studierenden, die einen einjährigen Deutsch-Vorbereitungskurs abschlossen, und Studierenden, die aus Deutschland zurückkehrten, aufdeckte. Kalkans Studie bestätigt keinen signifikanten Unterschied zwischen den Gruppen bei Korrekturen morphologischer Fehler (in Kasus und Deklination); unsere Forschung hingegen ergab statistisch signifikante Unterschiede bei morphosyntaktischen Fehlerkorrekturen.

5. Schlussfolgerung

Der Artikel befasst sich mit der Fehlerkorrektur von Deutschstudierenden pädagogischer Fakultäten in der Tschechischen Republik. Statistische Tests zeigten, dass die Anzahl der richtigen Korrekturen im deutschen Text mit Fehlern nicht vom Alter abhing, in dem die Studierenden mit dem Erlernen der deutschen Sprache begannen. Darüber hinaus gab es auch keinen statistisch signifikanten Unterschied in der Korrektur zwischen denjenigen Studierenden, die mit Deutsch vor Englisch begannen, und denen, die Englisch vor Deutsch lernten.

Tschechische Studierende waren im Vergleich zu MuttersprachlerInnen aus Deutschland und Österreich deutlich schlechter darin, Fehler zu korrigieren. Als wir die Fehlerkorrektur auf sprachlichen Ebenen betrachteten, fanden wir einen statistisch signifikanten Unterschied zwischen den beiden Studiengruppen bei morphosyntaktischen Fehlern, nicht jedoch bei lexikalischen und orthografischen Fehlern.

Bei der Analyse verwendete das Forschungsteam den parametrischen Shapiro-Wilk-Test und den nichtparametrischen Mann-Whitney-U-Test, der für den Vergleich zweier unabhängiger Datensätze geeignet ist, ohne dass eine Normalverteilung angenommen wird. Dieser Test ermöglichte es, Unterschiede zwischen den Gruppen in unseren linguistischen Daten genauer zu identifizieren. Außerdem wurde der Spearman-Korrelationskoeffizient angewandt.

Diese kombinierten statistischen Methoden erlauben es, genauere und zuverlässigere Ergebnisse aus Daten zu erhalten, die aus Korpusstudien stammen. Die Erweiterung dieser statistischen Methoden bereichert nicht nur die derzeitigen methodischen Ansätze in der Linguistik, sondern bietet auch eine solide Grundlage für künftige Forschungen. In Kombination mit bestehenden Studien und zukünftigen Forschungen trägt unser Ansatz zu einem tieferen Verständnis und innovativer Forschung in der Linguistik bei und bereichert damit theoretische und angewandte linguistische Studien (Baayen, 2008; Baroni & Evert, 2009; Biber & Jones, 2009).

Literaturverzeichnis

BAAYEN, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

BARONI, M. & EVERT, S. (2009). Statistical methods for corpus exploitation. *Corpus linguistics: An international handbook,* 2, 777–803. https://doi.org/10.1515/9783110213881.2.777.

BIBER, D. & JONES, J. K. (2009). Quantitative methods in corpus linguistics. *Corpus linguistics: An international handbook,* 2, 1286–1304. https://doi.org/10.1515/9783110213881.2.1286.
GRIES, S. T. (2009). *Quantitative corpus linguistics with R: A practical introduction*. Routledge.
KALKAN, H. K. (2021). Eine Studie zur fehleranalytischen Kompetenz der Studierenden in der Deutschlehrendenausbildung. https://www.researchgate.net/publication/348370924.
KÜHN, R. (1988). Muss man Korrigieren üben? Zu den Ergebnissen einer Erkundungsuntersuchung über das Korrekturverhalten ausländischer Deutschlehrer. *Deutsch als Fremdsprache*: Zeitschrift zur Theorie und Praxis des Deutschunterrichts für Ausländer, 25(1), 44–49.
ONDRÁKOVÁ, J. (2014). *Chyba a výuka cizích jazyků*. Gaudeamus.
ZHANG, Y. (2023). *A multidimensional analysis of language use in English argumentative essays: An evidence from comparable corpora.* SAGE Open*, 13*(3). https://doi.org/10.1177/21582440231197088.