

**KORPUSLINGUISTISCHE ANALYSE DER
FEHLERTYPEN BEI DER VERWENDUNG VON
“PRÄPOSITION+ARTIKEL+NOMEN” BEI DaF-
LERNENDEN**

**Almancayı Yabancı Dil Olarak Öğrenenlerin “Edat+Tanım
Edatı+İsim” Kullanımındaki Hata Biçimlerinin Derlem
Dilbilimi Kapsamında Analizi**

Miray ENEZ¹

Abstract

The present paper deals with the linguistic investigation of the data contained in a corpus received data with regard to error analysis. Firstly the relation between learner corpora and language acquisition research is represented; afterwards the learner corpus *Falko* is introduced; in the last point the learning texts of *Falko Essay L2v2.4* are examined with regard to error analysis. The aim of this paper is to show how goal-oriented is the use of prepositions with the right articles for advanced learners of the German as a foreign language. In this context the error annotation in the learner corpus *Falko* is used as an examination method. Finally, it can be seen that errors in article use have higher error rates than errors when using prepositions. In addition, the annotation of example texts shows that *tokens* are often assigned the *ZH1Diff-Tag CHA* (in context: false preposition or incorrect articles). With *frequency analysis* all types of errors and their frequency distribution were classified. Because the *target hypotheses* play an important role in this classification, we can say that *target hypotheses* or similar concepts are always required if linguistic structures cannot be described with given grammar. From the perspective of the grammatical level, we can conclude that the prepositions and articles are still an acquisition problem in the advanced field.

Keywords: Learner Corpora, Error analysis, German as a foreign language

Özet

Bu çalışma, bir derlemeye ait verilerin hata analizi kapsamındaki dilbilimsel incelemesi ile ilgilenmektedir. Bu bağlamda öncelikle öğrenci derlemeleri ile dil edinimi araştırmaları arasındaki ilişki ortaya konarak öğrenci derlemesi *Falko* tanıtılır; sonrasında *FalkoEssay L2v2.4* başlığında yer alan öğrenci metinleri, hata analizine dayanarak incelenir. Bu araştırmanın amacı, almancayı yabancı dil olarak ileri düzeyde öğrenen öğrencilerin, edat ve tanım edatı kullanımının erek dile ne kadar uygun olduğunu ortaya koymaktır. Bu kapsamda *Falko* derlemesindeki anotasyon, araştırma metodu olarak kullanılmıştır. Araştırmanın sonucunda tanım edatı kullanımındaki hata oranının, edat kullanımındaki hata oranına kıyasla daha yüksek olduğu tespit edilmiştir. Ayrıca hatalara ait örnek metinlerin anotasyonu, *örnekçelerin (tokens)* çoğunlukla *ZH1Diff-Tag CHA* başlığı altında (*yanlış edat veya yanlış tanım edatı kapsamında*) görüldüğünü ortaya koymaktadır. *Sıklık analizi* sayesinde tüm hata biçimleri ve bunların ne sıklıkta görüldüğü sınıflandırılmıştır. Bu sınıflandırmada *amaç hipotezleri* önemli bir rol oynadığı için, *amaç hipotezleri* veya benzer kavramların, dilsel yapıların varolan dilbilgisi kuralları ile tanımlanamadığı durumlarda oldukça

¹ Arş. Gör., Anadolu Üniversitesi mirayenez@anadolu.edu.tr

gerekli olduğunu söyleyebiliriz. Dilbilgisel açıdan yaklaşacak olursak, edat ve tanım edatı konularının ileri düzeyde yabancı dil kullanımında dahi sorun teşkil ettiği sonucuna varabiliriz.

Anahtar kelimeler: Öğrenci derlemeleri, hata analizi, yabancı dil olarak Almanca

Einführung

Anhand der Korpuslinguistik, die besonders durch die technologischen Entwicklungen in der Hard- und Softwaretechnik sowie der Sprachtechnologie in den letzten Jahrzehnten möglich wurde, sind Korpora zu wichtigen Hilfsmitteln in der linguistischen Forschung geworden. In diesem Zusammenhang entwickeln sich verschiedene Lernerkorpora zu einer wichtigen Ressource für eine empirisch fundierte Erfassung und Beschreibung von Lernersprache. Unter Lernerkorpora versteht man die Zusammenstellungen von Lernertexten, speziell von Fremdsprachenlernern (Nicht-Muttersprachlern):

»Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardized and homogeneous way and documented as to their origin and provenance« (Granger 2002: 7).

Im gleichen Zusammenhang erwähnt Granger weiter, dass der einzige Unterschied zwischen Lernerkorpora und Korpora nur darin liegt, dass die Daten bei Lernerkorpora aus den Fremdsprachenlernern stammen.

»Learner corpora have all the characteristics commonly attributed to corpora (...), the only difference being that the data come from language learners« (Granger 2008: 259).

Im Licht der aktuellen Forschungen betonen Ludeling/Walter (2009), dass in der Fremdsprachvermittlung und –erwerbsforschung die Verwendung von Korpora und Korpusdaten an immer mehr Bedeutung gewinnt. Mit den Korpora in der Sprachvermittlung gehen die Fokussierung vor allem auf den Sprachgebrauch und das Prinzip der Authentizität einher. Das Thema Authentizität wird häufig als problematisch gesehen und in Bezug auf Lernerkorpora viel diskutiert. Im Hinblick auf die Authentizität des Sprachmaterials weist Granger (2002: 8) daraufhin, dass die Daten in Lernerkorpora generell weniger authentisch sind als die in den meisten Korpora, die Daten von Muttersprachlern umfassen. Dies geht häufig davon aus, dass die Texte, die in der Klassenzimmersituation produziert werden, immer in irgendeiner Form nicht authentisch sind, da sie nicht in einer natürlichen und ungesteuerten Kommunikationssituation entstehen. Authentizität meint in der Lernerkorpuslinguistik, dass die Lerneräußerungen nicht allein für eine linguistische Untersuchung angefertigt wurden, sondern in einem natürlichen Kontext, und dies heißt, einem kommunikativen Kontext (vgl. Mukherjee 2002: 47).

In Korpora wie *Falko* wurden die Texte zwar in einem standardisierten Verfahren erhoben, die Lerner wussten aber nicht, dass sie Texte für ein Korpus verfassen. Aus der Perspektive der linguistischen Ebene liegt der Fokus meistens auf der Korrektheit der grammatischen Strukturen. In diesem Zusammenhang spielt die Authentizität in

² FL steht für „Foreign Language“; SL für „Second Language“; SLA für „Second Language Acquisition“ (Zweitspracherwerb), FLT für Foreign Language Teaching (Fremdsprachenvermittlung).

Lernerkorpora für geschriebene Sprache weniger Rolle als im Korpus gesprochener Sprache.

Diesen zwei Thematiken wird sich die folgende Arbeit widmen: Im ersten Schritt wird die Funktion vom Lernerkorpus *Falko* erläutert, wobei die Fehlerannotation berücksichtigt und die Relevanz und Äußerungsmerkmale von Fehler im Bedeutungsträger *Tags* beschrieben werden. Im Anschluss daran werden die Lernertexte im *Falko*-Korpus in Bezug auf Fehlertypen und ihre Häufigkeitsverteilung untersucht.

Seit dem Jahre 2004 entsteht in Kooperation zwischen dem Fachgebiet Deutsch als Fremdsprache an der Freien Universität und der Korpuslinguistik der Humboldt-Universität sowie unter Beteiligung der Georgetown University, Washington D.C. das frei zugängliche Lernerkorpus **Fehler**annotierte **Lerner**korpus (**Falko**). Das Korpus ist ohne vorherige Registrierung unter online <http://korpling.german.hu-berlin.de/falko/index.jsp> verfügbar. Dieses Korpus umfasst schriftliche Texte fortgeschrittener Lerner des Deutschen als Fremdsprache und die Texte von deutschen Muttersprachlern (Schmidt 2011: 562). Im Bereich *FalkoEssay L2*, der in dieser Arbeit untersucht wird, befinden sich Aufsätze, die von fortgeschrittenen Lernern des Deutschen erstellt wurden. Die Texte sind argumentative Aufsätze zu einem von vier vorgegebenen Themen, die aus der Gesamtmenge der im *International Corpus of Learner English* (ICLE) (Granger 1993: 2003) verwendeten Aufsatzthemen (Kriminalität, Entlohnung, Jugend und Studium) ausgewählt wurden. Die Lernertexte stammen von Nicht-Muttersprachlern, die teilweise an Feriensprachkursen an der Freien Universität Berlin und der Humboldt-Universität zu Berlin und teilweise an ausländischen Universitäten und Goethe-Instituten erhoben wurden.

Fehlerannotation

Im Allgemein kann *Annotation* folgendermaßen definiert werden:

„The practice of adding interpretative (especially linguistic) information to an existing corpus of spoken and/or written language by some kind of coding attached to, or interspersed with, the electronic representation of the language material itself“ (Granger 2002: 16).

Scherer (2006: 58) weist daraufhin, dass annotierte Korpora das Auffinden der für die Fragestellung benötigten Informationen erleichtern, weil sie nicht nur die Primärdaten, sondern auch Informationen über Grammatik, Struktur und Inhalt enthalten. Die Auszeichnung von Korpora durch Annotationen, d.h. zusätzlichen Merkmalen, wie z.B. die Wortartenzuweisung, ist hierbei von besonderem Interesse und Vorteil für Suchabfragen und linguistische Analysen.

Wie es in der Abbildung (1) zu sehen ist, lassen sich Korpora anhand einer Reihe von Kriterien klassifizieren. In Anlehnung an die Korпустypologie in Lemnitzer/Zinsmeister (2010: 103) ergibt sich die Annotation als Teil eines Korpus, während manche Autoren auf linguistische Annotation in Korpora ganz zu verzichten³. Dagegen behaupten Lemnitzer/Zinsmeister (2010), dass mehrere linguistische Ebenen annotiert werden können, wenn Annotation im Korpus vorhanden ist. Hier könnte man ggf. weiter

³ Lemnitzer/Zinsmeister (2010) haben die Annotation und Verfügbarkeit im Unterschied zu manchen dieser Autoren *„Sinclair (1996), Dodd (2000), Kenny (2000), Engelberg und Lemnitzer (2001), Atkins, Clear und Ostler (1992) als Kriterien aufgenommen.*

unterscheiden, z.B. in positionelle Merkmale, die dem einzelnen Token zugeordnet werden, wie die Morphosyntax, und in strukturelle Merkmale, die potenziell einen wortübergreifenden Charakter haben. Im Hinblick auf die Korpusuntersuchung in dieser Arbeit liegt der Fokus auf der Ebene von Fehlerannotation.

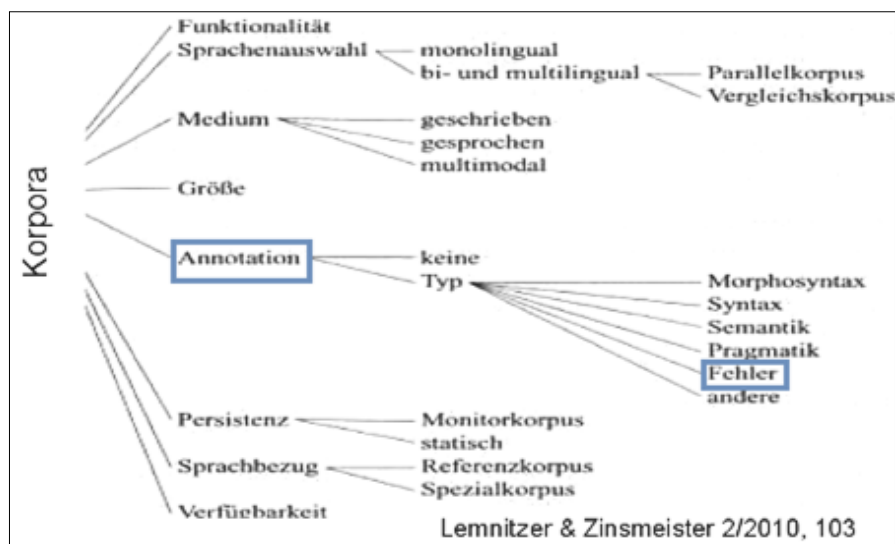


Abb.1. Korpustypologie: Übersicht über die Kriterien und ihre Werte

Bei der Auswertung von Lernerkorpora gibt es zwei grundlegende Strategien: Die erste Strategie basiert auf die Fehleranalyse *EA (Error Analysis)*, bei der Abweichungen von der Zielsprache untersucht werden. Die kontrastive Analyse bildet das Fundament für andere Strategie und beschäftigt sich mit dem Vergleich zielsprachlicher (grammatischer) Strukturen in L1 und L2 (Granger 2002: 12). Wie es in dieser Arbeit bearbeitet wurde, ist die Fehlerannotation eine wichtige Komponente der korpusorientierten Spracherwerbsforschung. Um einen Fehler annotieren zu können, muss man ihn identifizieren. Laut Corder (1986: 37) müssen dafür die Lerneräußerungen mit den „korrekt“ rekonstruierten Äußerungen verglichen werden.

Jede spezialisierte Lernerkorpora braucht für ihre Untersuchung spezielle Analysemethoden. Auch wenn die Annotationen wie *POS-Tagging* die Untersuchung von Lernerkorpora erleichtern, sollen sie durch andere Formen der Annotationen, die zu Analyse der Sprache von Lernern dienen, ergänzt werden (Granger 2002:18). Dazu gehört die Fehlerannotation, bei der je nach Annotationsschema die Fehler mit einem Tag versehen werden, der die Fehlerarten angibt (McEnery/ Xiao/ Yukio 2006: 42).

Ein Problem bei der Fehleranalyse ist die notwendige Annahme einer Zielhypothese – ein Fehler kann nur angenommen werden als eine Abweichung von einer (implizit angenommen oder explizit angegebenen) korrekten Form. Da es viele verschiedene solche korrekten Formen geben kann, kann dieselbe Struktur unterschiedlich analysiert werden. Dies kann zu unterschiedlichen Hypothesen über den Erwerb führen (Ludeling/Walter 2009 zit. nach. Tenfjord/Hagen/Johansen 2006). In diesem Rahmen erwähnen Ludeling/Walter (2009: 319) weiter, dass die Ergebnisse von Fehlerstudien ohne die zugrundeliegenden Daten und die Interpretationsgrundlage nicht einzuordnen

sind. Dafür ist es für die Nachvollziehbarkeit einer Analyse erforderlich, dass die Daten mit Zielhypothese und Annotation zugänglich sind.

Bei Lernerkorpus *Falko* erfolgt die Fehleranalyse durch die Fehlerkennzeichnung und -korrektur anhand der Zielhypothesen, die zur Analysemöglichkeit ungrammatischer Strukturen dienen. Dabei wird darauf geachtet, dass zwei widerstrebende Tendenzen während der Erarbeitung der Operationalisierung der *Falko*-Zielhypothese deutlich geworden sind:

»Je stärker man die Interpretation der Lerneräußerung lenkt und beschränkt, desto weniger Spielraum bleibt dem Annotatoren, die eigentliche Intention, des Lerners sinngemäß zu rekonstruieren. Je mehr von dieser Intention er jedoch versucht in der Zielhypothese widerzuspiegeln, desto weniger Einheitlichkeit verspricht das Resultat. Aus diesem Grund wurde für das Falko-Essay-Korpus entschieden, beide Strategien zu verfolgen, indem zwei verschiedene Zielhypothesen für den gleichen Text erstellt wurden.« (Reznicek et. al. 2012: 38).

In diesem Rahmen impliziert jede Fehleranalyse alternative Zielstruktur (Zielhypothese), welche eine Interpretation der Lerneräußerung durch einen geschulten Annotator darstellt. Zu jeder Lerneräußerung können gleichberechtigte, konkurrierende Zielhypothesen formuliert werden (vgl. Ludeling 2009). Wie es im Folgenden durch die Beispielanalyse illustriert wird, sind alle Strukturen anhand des Zielhypothesen-Ansatzes suchbar/trennbar, bei denen ein oder kein Grammatikalitätsproblem vorliegt. Um Fehler annotieren zu können wurde in dieser Arbeit *minimale Zielhypothese ZH1* verwendet. Der Grund hierfür liegt darin, dass durch die Operationalisierung der *ZH1* tokenbasierte Korrekturen möglich sind. Im Vergleich zu *erweiterter Zielhypothese ZH2* erlaubt die *minimale Zielhypothese ZH1* eine hohe Übereinstimmung zwischen den Annotatoren.

Ein weiterer Gewinn durch Zielhypothesen ist die Markierung der Fehler und ihre automatische und strukturelle Klassen (edit tags: CHA, DEL, MOVS, INS und SPLIT). Durch dieses Fehlertagset werden die Fehler zur Zielsprache in Beziehung gesetzt und auf dieser Basis interpretiert. Auf der Annotationsebene *DIFF* werden automatisch die Inhalte eines Tokens in Lernertext und *ZH1* verglichen und insgesamt fünf Tags vergeben. In der Abbildung (2) werden diese *ZH1Diff-Tags* und ihre Bedeutungen dargestellt. Diese Informationen kann man nutzen, um auf bestimmte Fehlertypen einzugrenzen oder die Häufigkeit der vorkommenden Fehlertypen festzustellen.

ZH1Diff-Tags	Bedeutung
CHA	"Token musste verändert werden"
DEL	"Token musste gelöscht werden"
MOVS	"Token musste bewegt werden"
INS	"Token musste eingefügt werden"
SPLIT	"Token musste geteilt werden"

Abb.2. Zur Bedeutung der ZH1Diff-Tags

Fehleranalyse der Beispieltexte

Fragestellung: Wo treten die Fehler bei der Verwendung von Präpositionen bzw. Artikeln bei fortgeschrittenen Lernenden des DaF auf?

Suche: Liste in der Zielhypothese eingefügten Präpositionen und Artikel.

Ausdruck in ANNIS Query Language: `pos="APPR" & pos="ART" & pos="NN" .1,2#2.#3 & #1 _=_ ZH1Diff`

Mit dem oben genannten Ausdruck erhält man alle Fälle, in denen die Präposition korrigiert werden musste (*ZH1Diff* taucht als zusätzliche Annotation unter der Präposition auf, wenn bei der Lerneräußerung dort etwas grammatisch nicht stimmt). Wie es in der Abbildung (3) zu sehen ist, taucht *CHA* als zusätzliche Annotation unter der richtigen Präposition *zu* auf, um darauf anzumerken, dass die Präposition im Lernertext verändert werden musste.

tok	.	Sie	gehen	dann	manchmal	an	der	Universität	.	um	eine	besuchen
ZH1	.	Sie	gehen	dann	manchmal	zu	der	Universität	.	um	eine	besuchen
ZH1DepID	58,000000	59,000000	60,000000	61,000000	62,000000	63,000000	64,000000	65,000000	66,000000	67,000000	68,000000	
ZH1Diff						CHA			INS			
ZH1S	s3	s4										
ZH1gpos	\$.	PPER	VVFIN	ADV	ADV	APPR	ART	NN	\$.	KOUI	ART	
ZH1gposDiff									INS	CHA		
ZH1lemma	.	Sie sie	gehen	dann	manchmal	zu	d	Universität	.	um	ein	
ZH1lemmaDiff						CHA			INS			
ZH1pos	\$.	PPER	VVFIN	ADV	ADV	APPR	ART	NN	\$.	KOUI	ART	
ZH1posDiff									INS	CHA		

Abb.3. Anwendungsbeispiel in FALCO in ANNIS (Falko Essay L2, cbs004_2006_09)

Wenn man alle Fehler an den Artikeln finden will, soll man in diesem Ausdruck `pos="APPR" & pos="ART" & pos="NN" .1,2#2.#3 & #1 _=_ ZH1Diff #1` durch `#2` austauschen. Wie es aus der Abbildung (4) ersichtlich ist, tritt *CHA* als zusätzliche Annotation unter den richtigen Artikeln *der* auf. Daneben ergibt sich ein anderes *ZH1Diff*-Tag *DEL* und das weist daraufhin, dass Token *von* im Lernertext gelöscht werden musste.

ZH1	Frauen	,	aber	eigentlich	auch		der	Männer)	zu	danken
ZH1DepID	97,000000	98,000000	99,000000	100,000000	101,000000		102,000000	103,000000	104,000000	105,000000	106,000000
ZH1Diff							DEL	CHA			
ZH1S	s4										
ZH1gpos	NN	,	KON	ADV	ADV		ART	NN	\$	PTKZU	VVINF
ZH1gposDiff							DEL				
ZH1lemma	Frau	,	aber	eigentlich	auch		d	Mann)	zu	danken
ZH1lemmaDiff							DEL				
ZH1pos	NN	,	KON	ADV	ADV		ART	NN	\$	PTKZU	VVINF
ZH1posDiff							DEL				
tok	Frauen	,	aber	eigentlich	auch	von	den	Männer)	zu	danken

Abb.4. Anwendungsbeispiel in FALKO in ANNIS (Falko Essay L2, fk004_2006_07)

Wenn man die Fehler bei der Verwendung von Präpositionen in der Kombination von *APPR+ART+NN* betrachtet, dann kann man erschließen, dass es dazu in Lernertexte in *Falko Essay L2v2.4* insgesamt 196 Treffer gibt. Dagegen ergeben sich anhand der Suchanfrage für die Fehler bei der Verwendung von Artikeln in der Kombination von *APPR+ART+NN* insgesamt 348 Treffer. Darüber hinaus kann man schlussfolgern, dass die Fehler bei der Artikelverwendung in den Lernertexten häufiger als die Fehler bei der Verwendung von Präpositionen sind.

Häufigkeitsliste

Eines der am meisten verwendeten Werkzeuge der Korpuslinguistik ist *frequency lists*. Es gibt verschiedene Arten von diesen Listen: Man kann z.B. lemmatisierte Daten oder nicht-lemmatisierte Daten verarbeiten. Weiterhin kann man sich n-gramme ansehen oder auch die Frequenz von POS(paaren) betrachten. Die häufigsten Wörter eines Korpus werden in der Computerlinguistik oft als *stopwords* bezeichnet. Solche *stopwords* sind (vermeintlich) nicht bedeutungstragend (vgl. Entrup 2014). Unter *Co-occurrence* bezeichnet man Kollokationen, die signifikant Häufiger auftreten, als es bei einer zufälligen Verteilung zu erwarten wäre. Ähnlich wie die Kollokation, handelt es sich bei der Konkordanz auch um ein Zusammentreffen von Wörtern. Dabei werden mehr als die direkten Nachbarn berücksichtigt. Die Menge der linken und rechten Nachbarn nennt man in der Computerlinguistik oft Fenster (window). Solche Analysen sind besonders interessant, wenn es z.B. darum geht zu ermitteln, welche Ergänzungen ein Verb bindet oder mit welchen Präpositionen verwendet wird (vgl. ebd.).

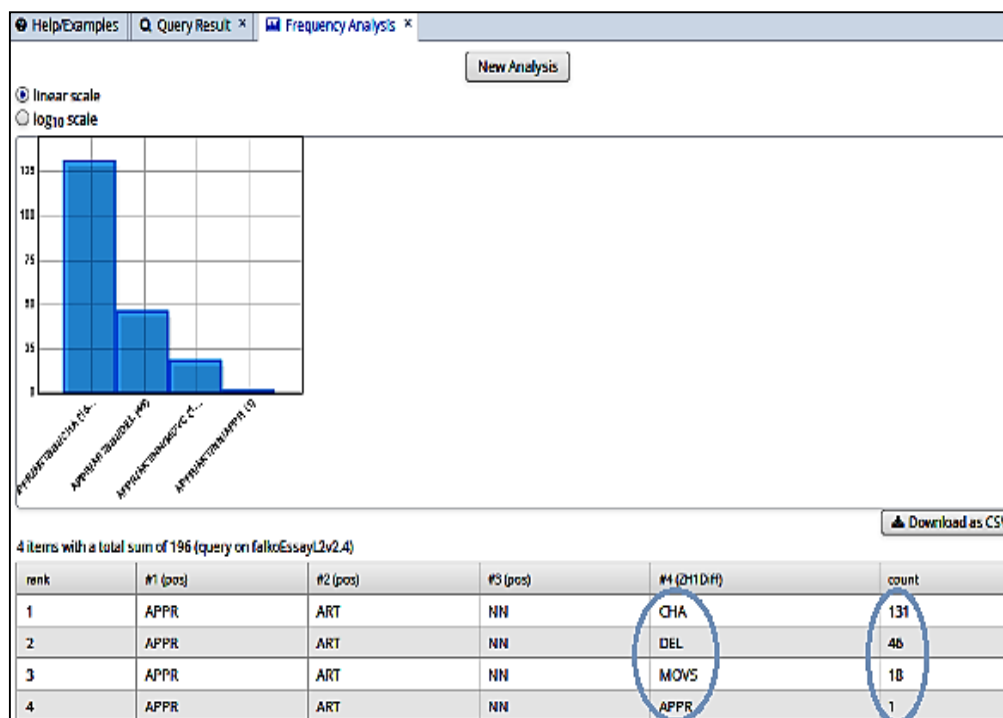


Abb. 5. ‚Frequency Analysis‘ in FALKO in ANNIS durch die Suchanfrage „pos=“APPR“ & pos=“ART“ & pos=“NN“ .1,2#2.#3 & #1 _=_ ZH1Diff“ (Falko Essay L2v2.4)

Wie es in der obigen Abbildung (5) dargestellt ist, ist es in *Falko* durch *frequency analysis* möglich, die Fehlertypen in den Lernertexten gemäß ihrer Häufigkeit zu klassifizieren. Aus der Suchanfrage *pos=“APPR“ & pos=“ART“ & pos=“NN“ .1,2#2.#3 & #1 _=_ ZH1Diff* ergibt sich die Anzahl der jeweiligen Fehlertypen, die von der Verwendung der Präpositionen ausgegangen sind. Laut der Häufigkeitsverteilung von Fehler tritt *ZH1Diff-Tag CHA* (Token musste verändert werden) mit 131 Treffern am häufigsten auf und weist daraufhin, dass es hier um falsche Präposition geht. An zweiter Stelle steht *DEL* (Token musste gelöscht werden) mit 46 Treffern und deutet an, dass die Präposition ungrammatisch ist. Danach kommt ein anderes *Diff-Tag MOVS* (Token musste bewegt werden) mit 18 Treffern. Anschließend ist *APPR* mit 1 Treffer zu sehen und es handelt sich hierbei um Fehler in den Daten, die da eigentlich nicht hingehören, aber in jedem Korpus vorkommen.

Im Vergleich zu Fehler, die bei der Verwendung von Präpositionen gesehen werden, tauchen die Fehler bei der Artikelverwendug mit 348 Treffern häufiger auf. Wie es in der Abbildung (6) abgebildet ist, gilt *ZH1Diff-Tag CHA* (Token musste verändert werden) als der am häufigsten vorkommenden Fehler mit 311 Treffern. Danach kommt *MOVS* (Token musste bewegt werden) mit 20 Treffern an zweiter Stelle, während dieser Fehlertyp bei der Verwendung von Präpositionen an dritter Stelle zu sehen ist. Dazu folgt *DEL* (Token musste gelöscht werden) mit 16 Treffern und als Letztes ist *ART* untergekommen, es geht hierbei wie bei *APPR* um dasselbe Problem, das in der Abbildung (5) zu sehen ist.

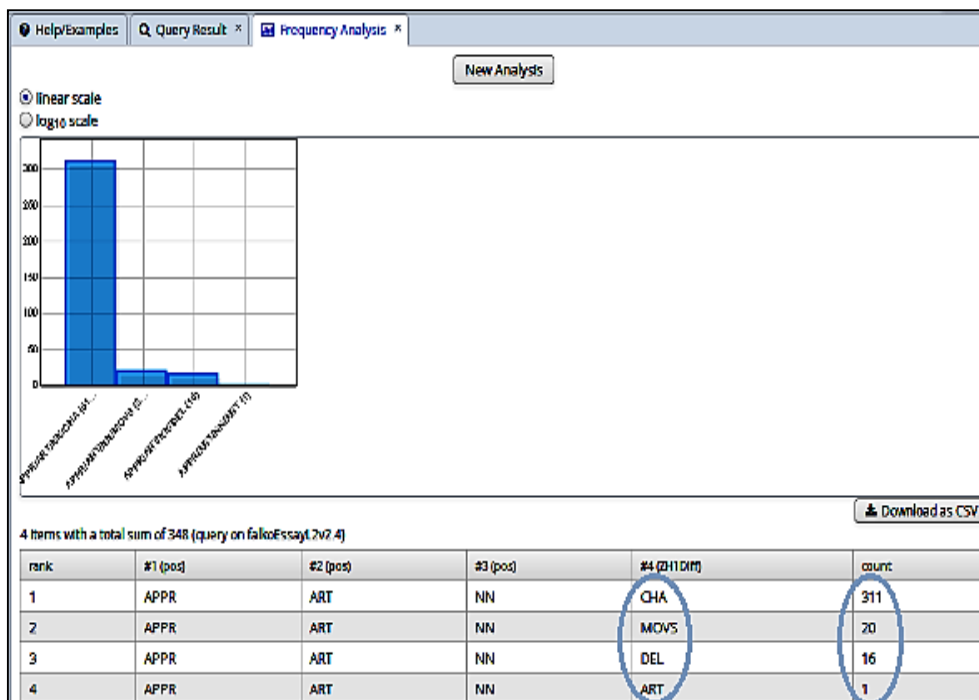


Abb.6. Frequency Analysis in FALKO in ANNIS durch die Suchanfrage „pos="APPR" & pos="ART" & pos="NN" & #1.1,2#2.#3 & #2 _ _ ZH1Diff"(Falko Essay L2v2.4)

META-Befehl

Wie es bisher erwähnt wurde, gibt es verschiedene Kriterien in Anlehnung an die Korpus Typologie in Lemnitzer & Zinsmeister (2010: 103). In Bezug auf die Sprachenauswahl aus dieser Typologie ergeben sich zwei Variante (die sog. monolingual und bi-/multilingual). Bei der Variante *bi-/multilingual* benutzt man häufig Vergleichskorpus zur kontrastiven Erforschung gesprochenen/geschriebenen Sprache. Dagegen ist es bei der Fehler-Annotation nicht möglich, die im Korpus enthaltenen Texte direkt zu vergleichen, aber man kann die Texte, die zur fehlerannotierten Korpora gehören, hinsichtlich verschiedener Aspekte klassifizieren. In den Metadaten der Essay-Texte in *Falko* findet man Ergebnispunkte von c-Tests, die im Rahmen der Erhebung durchgeführt wurden. Man kann versuchen, diese Informationen zu nutzen, um die Lernertexte bezüglich bestimmter Aspekte zu kategorisieren. Mit dem Zusatz & meta::ctest=/9./ erhält man beispielsweise alle Studierenden, die ein Testergebnis in den 90ern erzielt haben. Dieser Zusatz gibt nur Treffer von Probanden mit einem Ergebnis von 90-99 Punkten in den c-Tests, das in den Metadaten steht. Die genauen Metadaten pro Treffer erhält man, wenn er in der Trefferansicht auf den 'i'-Knopf drückt. & meta:: ist der Befehl für „Beachte bei der angegebenen Suche folgende Metadaten“. Wenn man beispielsweise & meta::sex="f" an seine Query anhängt, erhält man nur Treffer von Frauen. Aus der Abbildung (7) ist es ersichtlich, dass die Fehler von Frauen bei der Verwendung von Präpositionen (137 Treffer) häufiger als die Fehler (51 Treffer), die von Männern gemacht wurden, sind. Hieraus kann man sagen, dass es möglich ist, mit dem

Befehl `& meta::` einzelne Daten miteinander in Beziehung zu setzen und damit einen Vergleich zu machen.

The image shows two side-by-side screenshots of a Query Builder interface. Both screenshots display a search query: `pos="APPR" & pos="ART" & pos="NN" & #1.1,2#2.#3 & #1 _= ZH1Diff & meta::sex="f"` (left) and `pos="APPR" & pos="ART" & pos="NN" & #1.1,2#2.#3 & #1 _= ZH1Diff & meta::sex="m"` (right). Below the query input, there are buttons for 'Search', 'More', and 'History'. The left screenshot shows '137 matches in 83 documents (showing 6/10 results)', while the right shows '51 matches in 29 documents (showing 7/10 results)'. Both screenshots feature a 'Corpus List' section with a 'Visible' dropdown set to 'Falko' and a table of results.

Name	Texts	Tokens		
FalkoEssayL2v2.3	248	131.628	i	📄
falkoEssayL2v2.4	248	144.619	i	📄

Abb.7. Häufigkeitsverteilung der Fehler bei der Verwendung von Präpositionen nach Geschlecht

Fazit

Die vorliegenden theoretischen Erläuterungen sowie die betriebene Analyse zu Lernertexten haben gezeigt, wie fassbar und umsetzbar das Konzept der Fehleranalyse im Bereich der Korpuslinguistik ist. Ein adäquater Ersatz von fremdsprachlichen Texten durch rein onlinebasierte Lernerkorpora erscheint, gerade im Hinblick auf das übergeordnete Lernziel der korpuslinguistischen Untersuchungen, momentan noch als eine Wunschvorstellung. Das durch die vielfältigen und sich ständig weiterentwickelnden Technologien geprägte digitale Lernformat müsste sich verstärkt diesen technischen Bedingungen anpassen, um sich auf der einen Seite dem Potential des Lernerkorpus im Rahmen einer linguistischen bzw. traditionellen Untersuchung angleichen zu können, und um andererseits die Entfaltungs- und Vernetzungsmöglichkeiten des digitalen Raums zu einer Optimierung dieser konventionellen Formate im Kontext der digital vernetzten, globalisierten Gesellschaft auszunutzen (vgl. Rösler 2004: 78f).

Im Licht der aktuellen Forschungen haben Lernerkorpora wichtige Anwendungsgebiete in der Fremdsprachenvermittlung. Bei Lernerkorpora handelt es sich allerdings um spezielle Korpora, die neben allgemeinen Analysemethoden zusätzliche Techniken erfordern. Die große Schwierigkeit bei der Erforschung von Lernerkorpora liegt darin, dass sie fehlerhafte Sprache enthalten, die somit von den Sprachdaten, mit denen das Programm trainiert wurde, abweichen. Daher tauchen oft Schwierigkeiten, fehlerhafte Ausdrücke zu annotieren. Aus diesem Grund wurden in dieser Arbeit nur die Lernertexte aus *Falko Essay* annotiert und ein Anspruch auf die Fehlerannotation im Hinblick auf grammatische Ebene gestellt. Da die Summary-Texte in Falko nicht tiefer annotiert als

auf der oberflächlichen Lerneräußerung sind, ist es unmöglich, die Essays und Summaries technisch direkt zu vergleichen.

Bei der Fehlerannotation von Essay-Texten wurde geachtet, dass die Trennung zwischen zielsprachlicher Äußerungen und abweichender Äußerungen eine Zielstruktur bzw. Zielhypothese benötigt. In diesem Rahmen wurde das Verhältnis zwischen Lerneräußerung und Zielstruktur in abweichenden Fällen überprüft und die Ergebnisse durch die Anwendungsbeispiele illustriert.

Abschließend lässt sich feststellen, dass die Fehler bei der Artikelverwendung höhere Fehlerquote als die Fehler bei der Verwendung von Präpositionen aufweisen. Daneben zeigt die Annotation von Beispieltextrn bei den jeweiligen Fehlern, dass Tokens häufig das *ZH1Diff-Tag CHA (im Kontext: falsche Präposition oder falsche Artikel)* zugeordnet wird. Anhand *frequency analysis* wurden alle Fehlertypen und ihre Häufigkeitsverteilung klassifiziert. Da die Zielhypothesen dabei eine wichtige Rolle spielen, kann man sagen, dass Zielhypothesen oder ähnliche Konzepte immer erforderlich, wenn sprachliche Strukturen nicht mit gegebener Grammatik beschreibbar sind. Aus der Perspektive der grammatischen Ebene kann man zum Schluss kommen, dass die Präpositionen und Artikel auch noch im fortgeschrittenen Bereich ein Erwerbsproblem sind.

Reference

- Corder, S. P. (1986): The role of interpretation in the study. In: Corder, Stephen P. (Hrsg.): Error analysis and interlanguage. 4. impr. Oxford: Oxford University Press, S. 35–44.
- Entrup, B. (2014): Einführung in die Arbeit mit Korpora. Beschaffung von Daten und Umgang mit Korpora. Justus-Liebig-Universität Gießen.
- Granger, S. (2002): A Bird's-eye View of Computer Learner Corpus Research. In: Granger S., Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching (Language Learning and Language Teaching; 6). Amsterdam & Philadelphia; John Benjamins, S. 3-33.
- Granger, S. (2008): Learner corpora. In: Lüdeling, Anke; Kytö, Merja (Hg.): CorpusLinguistics. An International Handbook; 1. Berlin; de Gruyter, S. 259-275.
- Lemnitzer, L./ Zinsmeister, H. (2010.): Korpuslinguistik – Eine Einführung. Narr Studienbücher. 2. aktualisierte Auflage. Tübingen: Gunter Narr Verlag.
- Ludeling, A./ Walter, M. (2009): Korpuslinguistik. In: Christian Fandrych, Britta Hufeisen, Hans-J. Krumm & Claudia Riemer (Hrsg.): Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch. 2. überarbeitete Auflage. Mouton de Gruyter, Berlin, New York.
- McEnery, T./ Xiao, R./ Yukio, T. (2006): Corpus-based language studies. An advanced resource book, London: Routledge.
- Mukherjee, J. (2002): *Korpuslinguistik und Englischunterricht: Eine Einführung*. (Sprache im Kontext 14, ed. Ruth Wodak and Martin Stegu) Frankfurt am Main: Peter Lang.
- Reznicek, M. et al. (2012): Das Falko-Handbuch. Korpusaufbau und Annotationen Version 2.01. Tech. Rep. Humboldt-Universität zu Berlin.

Rösler, D. (2004): E-Learning Fremdsprachen. Eine kritische Einführung. Tübingen: Stauffenburg (=Stauffenburg-Einführungen 18).

Scherer, C. (2006): Korpuslinguistik. Heidelberg: Winter (=Kurze Einführung in die germanistische Linguistik 2).

Schmidt, K. (2011): Lernerkorpora: Ressourcen für die Deutsch-als-Fremdsprache-Forschung. In: Türkischer Internationaler Germanistik Kongress, 20.-22. Mai 2009. Tagungsbeiträge. Izmir, S. 555-573.

Online:

ANNIS Interface: <https://korpling.german.hu-berlin.de/falko-suche/> [24.03.2015]