Carolin Müller-Spitzer / Martina Nied Curcio /
María José Domínguez Vázquez / Idalete Maria Silva Dias /
Sascha Wolfer

Recherchepraxis bei der Verbesserung von Interferenzfehlern aus dem Italienischen, Portugiesischen und Spanischen: Eine explorative Beobachtungsstudie mit DaF-Lernenden

1	Einleitung	4.4	Auf der falschen oder richtigen
Spur: Die Bedeutung
der Ausgangshypothese
2	Beschreibung der Studie	5	Schluss
3	Datenaufbereitung	6	Literatur
4	Studienergebnisse		
4.1	Allgemeine Ergebnisse		
4.2	Suchstrategien		
4.3	Faktoren für die Korrektheit
der Verbesserung		

Abstract: In the past two decades, more and more dictionary usage studies have been published, but most of them deal with the question what users appreciate about dictionaries, which dictionaries they use and which information they need in specific situations. These studies presuppose that users indeed consult lexicographic resources. However, language teachers and lecturers of linguistics often have the impression that students use too few high-quality dictionaries in their every-day work. Against this background, we started an international cooperation project to collect empirical data evaluating that impression. Our aim was to evaluate what students (here from the Romance language area) actually do when they correct language problems. We used a new methodological setting to do this (screen recording

Carolin Müller-Spitzer, Institut für Deutsche Sprache, R 5, 6–13, 68161 Mannheim,
e-Mail: mueller-spitzer@ids-mannheim.de

Martina Nied Curcio, Università degli Studi Roma Tre, Via Ostiense, 236, 00146 Roma, Italien,
e-Mail: martina.nied@uniroma3.it

María José Domínguez Vázquez, Universidade de Santiago de Compostela, Avda. Castelao,
15782 Santiago de Compostela, Spanien, e-Mail: majo.dominguez@usc.es

Idalete Maria Silva Dias, Universidade do Minho Braga, Campus de Gualtar, 4710-057 Braga,
Portugal, e-Mail: idalete@ilch.uminho.pt

Sascha Wolfer, Institut für Deutsche Sprache, R 5, 6–13, 68161 Mannheim,
e-Mail: wolfer@ids-mannheim.de

<https://doi.org/10.1515/lexi-2018-0008>

with a thinking-aloud task). The empirical data we gained offers a broad insight into what language users really do when solving language-related tasks today.

Keywords: research into dictionary use, language learners, search strategies, observation study

Schlagwörter: Wörterbuchbenutzungsforschung, Sprachlernende, Suchstrategien, Beobachtungsstudie

1 Einleitung

Die empirische Wörterbuchbenutzungsforschung hat in den letzten 10–15 Jahren erhebliche Fortschritte erzielt (vgl. u. a. Töpel 2014, Welker 2014, Lew 2011, Lew 2015a). Auch die am Institut für Deutsche Sprache (IDS) durchgeführten Studien zu Onlinewörterbüchern (vgl. Müller-Spitzer 2014) haben den Kenntnisstand des Forschungsgebietes erweitert (vgl. Lew 2015b: 251). Doch fast alle Studien aus dem Bereich der Wörterbuchbenutzungsforschung beschäftigen sich mit der Frage, was Nutzerinnen und Nutzer an Wörterbüchern schätzen (u. a. Domínguez Vázquez/Mirazo Balsa/Vidal Perez 2013, Domínguez Vázquez/Valcárcel Riveiro 2015, Müller-Spitzer/Koplenig 2014), welche Wörterbücher bzw. welche Angaben sie in welchen Situationen benötigen (u. a. Koplenig/Müller-Spitzer 2014, Nied Curcio 2013), welche Präsentationsformen am besten geeignet sind (u. a. Lew 2010, Lew/Grzelak/Leszkowicz 2013), oder welche Angaben in Onlinewörterbüchern am meisten nachgeschlagen werden (u. a. Hult 2012, Koplenig/Meyer/Müller-Spitzer 2014). Diese Studien setzen damit immer voraus, dass lexikografische Hilfsmittel tatsächlich benutzt werden bzw. versetzen die Probandinnen bzw. Probanden in eine konkrete Situation, in der sie sich vorstellen, lexikografische Hilfsmittel zu nutzen. Gleichzeitig haben viele Sprachlehrende und Dozenten der Sprachwissenschaften den Eindruck, dass in der alltäglichen Realität von Studierenden zu wenig (gute) Wörterbücher genutzt werden (vgl. z. B. Frankenberg-Garcia 2011). Dementsprechend gibt es eine Lücke zwischen dem Feld, in dem die empirische Wörterbuchbenutzung sich bewegt, und den tatsächlichen alltäglichen sprachlichen Herausforderungen von Sprachlernenden oder Studierenden: Uns stehen noch zu wenige empirische Daten zur Verfügung, um beurteilen zu können, welche Rolle Wörterbücher heute z. B. in der alltäglichen Arbeit von Sprachlernenden spielen. Dies sehen andere Forscher/innen ähnlich, wie z. B. Frankenberg-Garcia 2011 oder Levy/Steel 2015: 194:

The study reported here, with data drawn from a large-scale survey, reports on what students say they do when using electronic dictionaries. This reportage does not necessarily reflect what students actually do [...]. Smaller-scale studies are needed to complement and enrich the findings of the present study.

Eine zentrale Aufgabe bei der Weiterentwicklung der empirischen Wörterbuchbenutzungsforschung ist es daher, potentielle Benutzerinnen und Benutzer bei der tatsächlichen Lösung ihrer sprachlichen Aufgaben, bei denen lexikografische Daten eine Rolle spielen, zu beobachten. So können wir uns empirisch der Frage nähern, wie oder ob bestimmte Gruppen heute überhaupt noch Wörterbücher nutzen und ob sie sie bewusst von anderen sprachbezogenen Daten im Web unterscheiden. Dafür müssen wir einen Versuchsaufbau schaffen, der nicht die Benutzung bestimmter Nachschlagewerke vorschreibt, sondern versucht, die Versuchssituation möglichst eng an ein alltägliches Umfeld anzunähern und die Recherchepraxis mit möglichst wenigen Vorannahmen zu beobachten. Eine solche empirische Untersuchung ist aufwändig. Sie könnte aber gerade in dieser Zeit, in der sich die Lexikografie „at a turning point in its history“ (Granger 2012: 10) befindet, sehr interessante Daten liefern, wie die Rolle von Wörterbüchern für das Erlernen und die Vermittlung von Sprache einzuschätzen ist. Wenn solche Daten in größerer Breite erhoben werden, können sie Hinweise darauf geben, inwiefern die Rolle von Wörterbüchern, z. B. durch mehr Wörterbuchdidaktik im Sprachunterricht, gestärkt werden könnte.

Da die empirische Wörterbuchbenutzungsforschung am IDS im Projekt „Empirische Methoden“ eine langfristige Projektaufgabe ist, sehen wir es auch als unsere Aufgabe an, das Forschungsfeld im oben skizzierten Sinne zu erweitern. Wir haben bereits in einer Pilotstudie an der Uni Mannheim mit muttersprachlichen Studierenden eine neue Form der Datenerhebung, eine Art „Massenlaborstudie“, getestet. Diese Studie hat interessante Daten dazu geliefert, welche Effekte lexikografische Ressourcen auf die Qualität der Textüberarbeitung haben, z. B. dass jene Teilnehmerinnen und Teilnehmer, denen lexikografische Hilfsmittel zur Verfügung standen, Texte besser optimieren konnten. Darüber hinaus konnten wir belegen, dass die Textoptimierung auch hinsichtlich einer Kosten-Nutzen-Relation effektiver ist, wenn lexikografische Ressourcen vorliegen (vgl. Wolfer et al. 2015).

Mit der hier beschriebenen Studie sind wir noch einen Schritt weitergegangen: Zum einen haben wir eine zentrale Benutzergruppe für Wörterbücher – die Sprachlernenden – in den Blick genommen. Dies war nur möglich durch die enge Zusammenarbeit mit Kooperationspartnerinnen aus dem Ausland (die Mitautorinnen dieses Beitrags), die sich im Gebiet „Deutsch als Fremdsprache (DaF)“ wesentlich besser auskennen als wir. Zum anderen haben wir methodisch das Paradigma einer (weitestgehend) ungesteuerten Beobachtungsstudie eingesetzt, um noch näher an das Verhalten im Sprachlernalltag heranzukommen und empirische Erkenntnisse hierzu zu gewinnen. Empirische Daten dieser Art können unserer Meinung nach langfristig eine Grundlage dafür schaffen, Wörterbuchdidaktik besser in den Sprachunterricht zu integrieren. Denn im Moment weiß man nicht, wo genau dieser Unterricht hinsichtlich der Wörterbuchbenutzung ansetzen soll: Wie nutzen DaF-Lernende jetzt lexikografische Ressourcen? Welche Suchstrategien wenden sie an? Unterscheiden sie zwischen unterschiedlichen Arten von Ressourcen? Welche Strategien erweisen sich als besonders erfolgreich? Das Ziel dieser Studie ist dem-

nach, empirische Daten zu sammeln, um besser zu wissen, wie DaF-Lernende tatsächlich hinsichtlich der Nutzung von Online-Sprachressourcen arbeiten. Selbstverständlich bietet unsere Studie noch keine umfassende Antwort auf diese Fragen, aber sie zeigt, dass sich auf diesem Weg wertvolle Erkenntnisse gewinnen lassen.

Unser Beitrag ist folgendermaßen gegliedert: Im zweiten Abschnitt beschreiben wir den Aufbau der Studie. In Abschnitt 3 erläutern wir, wie die durch Screenrecording und Thinking-Aloud gewonnenen Daten aufbereitet wurden, um sie quantitativ auswerten zu können. Der Hauptteil dieses Artikels – Abschnitt 4 – enthält die Ergebnisse der Studie. Wir berichten zunächst allgemeine Ergebnisse (4.1), werden dann näher auf die Suchstrategien der Studierenden eingehen (4.2) und Faktoren für die Korrektheit der Verbesserungen herausarbeiten (4.3). Wir beenden den Ergebnisteil mit einer kurzen Ausführung, wie wichtig die Ausgangshypothese der Studierenden zu sein scheint (4.4) und ziehen abschließend ein Fazit.¹

2 Beschreibung der Studie

Wie in der Einleitung erwähnt, haben wir unsere Studie auf DaF-Lernende aus dem romanischen Sprachraum beschränkt, um das gleiche Studiendesign an allen Erhebungsorten einsetzen zu können. Aus diesem Grund haben wir uns auf Interferenzfehler konzentriert, die für DaF-Lernende aus Italien, Portugal und Spanien gleichermaßen zutreffen, sodass wir die gleichen fehlerhaften Sätze allen Studierenden zur Verbesserung vorlegen konnten. Unsere Probandinnen und Probanden waren, orientiert am Gemeinsamen Europäischen Referenzrahmen für Sprachen, etwa auf dem A2/B1-Niveau und hatten alle eine ähnliche Anzahl von Deutschstunden an der Universität absolviert. Einen Sprachtest vor der Studie haben wir allerdings nicht durchgeführt. Wir werden am Schluss darauf eingehen, warum wir das für zukünftige Studien anders machen würden. Die Erhebung wurde an drei Orten durchgeführt: In Braga in Portugal, Santiago de Compostela in Spanien und Rom in Italien. Pro Ort haben wir im März/April 2017 Daten von 14 Probandinnen und Probanden erhoben. Jede/r Teilnehmende hat 18 deutsche Sätze mit einem Inter-

¹ Eine englischsprachige Publikation zu dieser Studie ist in Arbeit. Voraussichtlicher Titel: „The right hypotheses and careful reading make the difference: Results of an observation study with language learners on using language resources online“ (gleiche AutorInnen wie dieser Beitrag). Außerdem ist in den diesjährigen (2018) Euralex-Proceedings ein eher methodisch ausgerichteter Beitrag zu dieser Studie bereits erschienen, mit dem Titel: „Combining Quantitative and Qualitative Methods in a Study on Dictionary Use“. Wir danken außerdem Alexander Koplenig aus dem IDS-Team „Empirische Methoden“ für die Diskussion der Studienergebnisse, allen Hilfskräften und Werkvertragsnehmer/innen, die an der Studie beteiligt waren sowie den Teilnehmerinnen und Teilnehmern des IDS-Kolloquiums im Herbst 2017, des EMLex-Kolloquiums in Stellenbosch sowie der FaDaF-Tagung 2018 in Mannheim, mit denen wir die Studienergebnisse diskutiert haben. Ein besonderer Dank geht an die Teilnehmenden der Studie für ihre Mitarbeit.

Tab. 1: Stimulus-Sätze der Studie.

ID	Satz
1	Meine Nachbarin möchte immer alles wissen. Sie ist sehr kurios .
2	Leider kann ich heute nicht Tennis spielen. Ich bin zu besetzt .
3	Bist du bereit ? Wir müssen jetzt los, wir sind sowieso schon zu spät dran.
4	Kein Problem, wenn der Zucker beendet ist; ich nehme dann Honig.
7	Ich bin einverstanden mit dir .
9	Das erlaube ich dir nicht. Es ist außer Frage .
11	An unserem Forschungsinstitut ist Ihnen unsere Bibliothek 24 Stunden zur Verfügung .
12	Obwohl ich studiere, wohne ich noch mit meinen Eltern.
14	Wenn ich zur Schule ging, habe ich viel Sport gemacht.
15	Morgen habe ich einen Arzttermin und kann deshalb nicht in die Klasse gehen .
18	Ich vorbereite gerade meine letzte Prüfung.
19	Ich möchte ein Stipendium beim DAAD bewerben .
20	Ich habe die Hose viel zu klein gekauft. Jetzt muss ich nochmals ins Geschäft zurück und sie wechseln .
21	Obwohl sich der Junge beeilt hat, hat er die U-Bahn verloren .
22	Er wohnt seit Jahren in Berlin und trotzdem verliert er sich immer noch.
24	Um beim Kartenspielen zu gewinnen, musst du exakt die Regeln folgen .
25	Der Artikel handelt sich um die Migranten in Deutschland.
26	Ich möchte dir heute über einen interessanten Artikel sprechen .

ferenzfehler erhalten. Die Fehler wurden so konstruiert, dass sie drei Anforderungen erfüllen: i) Der Fehler ist typisch für relativ frühe Deutschlernende mit einer Muttersprache aus den romanischen Sprachen. ii) Der Fehler kann nicht durch einfaches Suchen im Web nach dem Stimulus-Satz oder dem Teil des Satzes, der den Fehler enthält, behoben werden. iii) Der Fehler muss möglichst so schwer zu beheben sein, dass die Studierenden dafür Hilfsmittel benötigen und die Aufgabe nicht aus ihrer eigenen Kompetenz heraus lösen können. Diesen Anforderungen gerecht zu werden, war eine größere Herausforderung als wir zunächst dachten, da viele der Standard-Interferenzfehler durch eine Google-Recherche gelöst werden konnten. Durch mehrere Pretests abgesichert konnten dann jedoch 18 Sätze ausgewählt werden, in denen die Fehler jeweils markiert waren (vgl. Tabelle 1).

Der Ablauf der Studie hat jeweils folgendermaßen ausgesehen: Den Studierenden wurde im Vorfeld gesagt, dass sie an einer Studie teilnehmen sollten, die allerdings keinerlei Einfluss auf ihre Kursbewertung haben würde. In den Pretests hatte sich gezeigt, dass es wichtig ist, dass bei der Durchführung der Studie keine Lehrperson im Raum ist, andernfalls haben die Studierenden zu oft gefragt, ob dies und jenes wirklich erlaubt sei. Nach dem Unterschreiben der Einverständniserklärung in der jeweiligen Landessprache wurde den Probandinnen und Probanden gesagt, dass sie nun fehlerhafte deutsche Sätze vorgelegt bekommen und diese möglichst so bearbeiten und verbessern sollen, wie sie dies auch zuhause tun würden. Sie könnten also frei im Internet recherchieren und die Ressourcen nutzen,

die sie wollen. Gleichzeitig sollten sie laut denken, also kommentieren, ob sie den Fehler erkennen, warum sie welche Ressourcen aufrufen, ob sie zufrieden mit den Ergebnissen sind etc. Diese Instruktionen wurden den Teilnehmenden auch schriftlich in ihrer Muttersprache vorgelegt. Sie haben dann an einem von uns gestellten Laptop gearbeitet, damit die technischen Bedingungen für alle gleich waren. Vorinstalliert waren als Browser Google Chrome und Mozilla Firefox, als Sprache wurde die jeweilige Landessprache eingestellt und der Browsercache und -verlauf wurden nach jeder Versuchsperson gelöscht, um Reihenfolgeeffekte zu vermeiden. Zusätzlich zu einem von uns (Carolin Müller-Spitzer oder Sascha Wolfer) als Versuchsleiter/in war eine Person anwesend, die die Landessprache beherrschte, damit die Studierenden in ihrer Muttersprache Fragen stellen konnten. Diese zusätzliche Person hatte darüber hinaus die Aufgabe, die Studierenden regelmäßig daran zu erinnern, dass sie laut denken sollten (natürlich in ihrer Muttersprache). Keine dieser Personen war so im Raum platziert, dass man auf den Laptop-Bildschirm schauen konnte. So sollte trotz der doch etwas künstlichen Atmosphäre eine möglichst geschützte Situation für die Probandinnen und Probanden geschaffen werden. Als Screenrecording-Software nutzten wir die kostenlose Version der Software *Active Presenter*. Die Audiospur wurde über ein externes Mikrofon aufgenommen und nachträglich als Tonspur in den Bildschirmfilm integriert, um eine synchrone Untersuchung von Bildschirmaufnahmen und Think-Aloud-Protokollen (im Folgenden: TAPs) zu ermöglichen. Die Studierenden hatten jeweils 45 Minuten Zeit, die Stimulus-Sätze zu verbessern. Dieser Zeitrahmen war ihnen allerdings nicht bekannt, da wir nicht wollten, dass sie sich von Anfang an zu sehr beeilen. Nach 30 Minuten haben wir sie allerdings darauf hingewiesen, dass nun noch 15 Minuten verblieben. Diese zeitlichen Restriktionen mussten wir vorgeben, um die Erhebung zeitlich planen zu können. Im Ergebnis haben wir pro Proband/in ca. 45 Minuten Bildschirmfilm erhalten. Dies sind in der Summe für die Studie etwa 1935 Minuten Filmmaterial plus TAPs mit mehr als 2.200 Suchaktionen. Um diese Daten sinnvoll auswerten zu können, war ein relativ aufwändiger Prozess der Datenaufbereitung notwendig, der im nächsten Abschnitt beschrieben wird.

3 Datenaufbereitung

Um statistische Auswertungen zu ermöglichen, mussten sowohl die Bildschirmfilme, die TAPs sowie die Lösungen der Studierenden aufbereitet werden. Die Bildschirmfilme wurden in Excel-Tabellen überführt, in denen eine Vielzahl von Spalten angelegt wurden, um verschiedenste Auswertungen zu ermöglichen wie den Namen der aufgerufenen Ressourcen, ob diese gezielt aufgerufen wurden, den genauen Suchstring oder die Sprache des Suchstrings (ein Auszug dieser Auflistung findet sich in Tab. 2). Der Aufwand für diese Überführung der Bildschirmfilme in auswertbare Daten betrug ca. 210 HiWi-Stunden.

Tab. 2: Auszug aus den Auswertungskategorien für die Überführung der Bildschirmfilme in statistisch auswertbare Daten.

ID Prob	Aus dem Excel-Dokument mit den Sätzen zu entnehmen (z. B. „B-01“ oder „S-03“)
Erhebungsort	B/S/R, ableitbar aus der ID
Aktueller Satz ID	In der linken Spalte des Excel-Dokuments
Aktueller Satz POS	Position des Satzes für den/die TeilnehmerIn, also die Reihenfolge im Excel-Dokument des einzelnen Probanden (die Sätze wurden für alle Studierenden randomisiert)
Gesamtzeit	Gesamtzeit des Films
Timestamp	Zeitanzeige im Bildschirmfilm, an dem eine Aktion ausgeführt wurde
Name Ressource	Name der Ressource, die aufgerufen wurde (z. B. google.es, pons.eu, DWDS, duden.de usw.) bzw. Excel (für die Zeit, die die ProbandInnen im Excel-Dokument arbeiten)
Ressource gezielt aufgerufen (j n)	Auf „j“ gesetzt, wenn eine Ressource gezielt aufgerufen wurde, ohne dass ein Suchstring in einer Suchmaschine eingegeben wurde, also nur wenn direkt z. B. www.duden.de aufgerufen wurde und <i>nicht</i> , wenn ein Wort (z. B. „bereit“) gesucht wurde und dann einer der Treffer (z. B. duden.de) aufgerufen wurde.
Suchstring	Genauer Suchstring, den der/die TeilnehmerIn eingegeben hat
Sprache Suchstring	(de=Deutsch en=Englisch es=Spanisch it=Italienisch pt=Portugiesisch andere)
Aktion	Kurze Erläuterung wie: ruft Duden.de auf, geht wieder auf den bereits geöffneten Tab mit dwds.de etc.

Die TAPs wurden in die jeweilige Landessprache transkribiert und ins Deutsche übersetzt. Der Aufwand hierfür war mit ca. 350 HiWi-Stunden sehr hoch. Diese Daten wurden anschließend in die Tabellen zu den Bildschirmfilmen integriert. Zusätzlich wurde auf Basis der TAPs vermerkt, wann wir genau die Zeitwarnung gegeben haben (dass noch 15 Minuten Zeit zur Verfügung stünden), damit wir später auswerten konnten, ob sich das Rechercheverhalten danach systematisch verändert hat.

In einem dritten Schritt wurden die verbesserten Sätze der Proband/innen dahingehend annotiert, ob sie richtig, falsch oder gar nicht verbessert wurden. Diese Annotation wurde zu zweit unabhängig voneinander durchgeführt. Im Anschluss wurden diese Annotationen automatisch verglichen. Das Inter-Annotator-Agreement bei den insgesamt 816 Sätzen war sehr gut: $\kappa_{\text{mean}} = 0.82$. Die unterschiedlich annotierten Sätze wurden dann mündlich durchgesprochen und klar eingeordnet. Der Aufwand für die Annotation insgesamt war mit ca. zehn Arbeitsstunden sehr gering.

Auch die Kategorisierung der Ergebnissätze wurde dann in die Gesamtauswertungstabelle überführt, sodass eine Gesamttabelle zur Auswertung der Studie vorlag. Alle weiteren statistischen Auswertungen wurden mit R^2 durchgeführt.

4 Studienergebnisse

4.1 Allgemeine Ergebnisse

Wie bereits in Abschnitt 2 beschrieben, konnte jede/r Proband/in 18 Sätze korrigieren.³ Durchschnittlich wurden davon 16 Sätze bearbeitet. Diese Zahl ist über die drei Erhebungsorte sehr ausgewogen (vgl. Abb. 1 links). Die Zahl der korrekt bearbeiteten Sätze unterscheidet sich allerdings erheblich: Während in Rom und Santiago de Compostela etwa die Hälfte der Sätze korrekt verbessert wurden, sind es in Braga kaum mehr als 15 % (vgl. Abb. 1 rechts). Dies deutet darauf hin, dass das Sprachniveau der Probandinnen und Probanden aus Braga deutlich niedriger war als das Sprachniveau derjenigen an den anderen zwei Orten. Da wir allerdings keinen Sprachtest vor der Studie durchgeführt haben, können wir nur vermuten, dass genau das die Ursache für die starken Unterschiede ist.

Abb. 1: Anzahl bearbeiteter (links) und korrekter Sätze (rechts).

² R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

³ Aus Versehen bekamen die ersten ProbandInnen in Braga mehr Sätze vorgelegt (aus einer vorhergehenden, umfangreicheren Version, die aufgrund der Erfahrung aus den Pretests reduziert wurde). Der Irrtum wurde allerdings nach den ersten drei ProbandInnen bemerkt.

Abb. 2: Korrektheit der einzelnen Sätze.

Es zeigt sich auch, dass die einzelnen Sätze für die Teilnehmerinnen und Teilnehmer ein sehr unterschiedliches Schwierigkeitsniveau hatten (vgl. Abb. 2). Während z. B. der Satz „Kein Problem, wenn der Zucker **beendet** ist; ich nehme dann Honig“ (Satz Nr. 4) in nur etwa 10 % der Bearbeitung korrekt verbessert ist, sind über 60 % der Verbesserungen für den Satz „Leider kann ich heute nicht Tennis spielen. Ich bin zu **besetzt**.“ (Satz Nr. 2) korrekt. Ein klarer Zusammenhang zwischen Fehlertyp und Verbesserungsrate lässt sich allerdings nicht erkennen. Vielmehr scheint es eine Vielzahl von Faktoren zu sein, die Sätze schwer bzw. leicht zu verbessern machte, u. a. wie gut die Studierenden es schafften, den markierten Fehler in eine sinnvolle Online-Recherche zu transformieren. In Abschnitt 4.3 wird näher darauf eingegangen, welche Faktoren wir als einflussreich für die korrekte Verbesserung identifiziert haben.

Für die Anzahl der Suchvorgänge pro Satz zeigt sich: In 108 Bearbeitungen wurde nur eine Suche durchgeführt, in 233 Bearbeitungen zwei bis drei Suchen und in 246 Fällen mehr als vier (vgl. Abb. 3).

Bei der Überführung der Bildschirmfilme in auswertbare Daten wurde bei einer Suche über Google jeweils markiert, an welcher Position der Suchergebnisliste der aufgerufene Treffer war. So können wir auswerten, wie oft welche Position der Google-Suchergebnisliste aufgerufen wurde. Das Ergebnis ist in Abbildung 4 zu sehen: In nennenswerter Anzahl werden nur die ersten vier Treffer der Suchergebnisliste aufgerufen (d. h., fast niemand hat gescrollt, denn 4 bis 5 Ergebnisse waren auf dem Laptopbildschirm direkt zu sehen, je nachdem, ob das Fenster für den Google Translator eingeblendet war oder nicht). Von diesen haben sich aber mehr als die Hälfte der ProbandInnen (63 %) auf den ersten Treffer konzentriert.

Wie anfangs erwähnt, wissen wir wenig darüber, welche Typen von Ressourcen Studierende tatsächlich nutzen, wenn sie sprachliche Probleme lösen. Manche

Abb. 3: Anzahl Suchvorgänge pro Satz.

Abb. 4: Position aufgerufenes Google Suchergebnis.

Lehrende behaupten, dass Studierende heutzutage immer mehr automatische Übersetzungsprogramme nutzen und kaum mehr Wörterbücher.⁴ Nach Selbstauskunft werden wiederum sehr häufig Onlinewörterbücher genutzt (Levy/Steel 2015: 9, Koplenig/Müller-Spitzer 2014: 130). Eine wichtige Frage in unserer Studie

⁴ So z. B. die mündliche Auskunft zweier Lehrkräfte beim *Berlitz Language Center* in Mannheim und auch einiger Anwesenden bei der Tagung des *Fachverbands für Deutsch als Fremd- und Zweitsprache (FaDaF)*, auf dessen Jahrestagung (2018) wir die Studienergebnisse vorgestellt haben.

Tab. 3: Ressourcen, die in Braga, Santiago de Compostela und Rom benutzt wurden mit Angaben, wieviel Prozent der Teilnehmer/innen sie verwendet haben und in wie vielen Satzbearbeitungen sie anteilig verwendet wurden.

Ressource	% Teilnehmer	% Satzüberarbeitungen
Google	100,0 %	38,6 %
Pons Dictionary	79,1 %	54,6 %
Pons Translator	37,2 %	11,0 %
Reverso Conjugation	34,9 %	7,29 %
Linguee	32,6 %	11,0 %
Google Translator	27,9 %	16,3 %
Verbformen	27,9 %	7,12 %
Duden	25,6 %	7,12 %
Bab,la Dictionary	16,3 %	5,59 %
Dict.cc	14,0 %	1,19 %
Deutsch Info	14,0 %	3,39 %
Lingolia Deutsch	11,6 %	1,86 %
WordReference	9,3 %	2,71 %

war daher für uns, erstens zu erfahren, welche Arten von Ressourcen unsere Probandinnen und Probanden nutzen und zweitens zu sehen, ob sie unterschiedliche Suchstrategien für verschiedene Typen anwenden.

Zunächst zeigt sich, dass die Studierenden in unserer Studie eine Vielzahl von Ressourcen und vor allem auch eine Vielzahl unterschiedlicher Typen nutzten. Tabelle 3 zeigt die Liste der Quellen, die an allen drei Orten verwendet wurden. Schon an der Liste zeigt sich, dass man mit Studierenden genau die unterschiedlichen Ressourcen im Unterricht besprechen sollte: Beispielsweise waren die an der Studie beteiligten Dozentinnen überrascht, dass so viele Studierende den Pons-Translator verwenden. Anscheinend, so kann man zumindest vermuten, haben sich einige Studierende gemerkt, dass „Pons“ eine ‚gute‘ Seite ist. Dort verwenden sie aber nicht nur das Wörterbuch, sondern eben auch das Übersetzungsprogramm, welches allerdings zumindest für die hier untersuchten Sprachen wesentlich schlechter funktioniert als „Google Translate“ („DeepL“ war zum Zeitpunkt der Studierendurchführung noch nicht auf dem Markt).

Von den Typen der Ressourcen her (Abb. 5) zeigt sich bei unseren Probandinnen und Probanden nicht, dass weniger Wörterbücher als Übersetzungsprogramme genutzt werden: Am meisten werden Wörterbücher oder Wörterbücher mit Grammatiktabellen genutzt, gefolgt von Suchmaschinen (die natürlich auch benutzt werden, um Ressourcen aufzurufen, in dem z. B. „Duden online“ in Google eingegeben wurde). Zwar sind 42 Probandinnen und Probanden keine große Anzahl, dafür sind die Daten aber in dem Sinne verlässlich, dass wir die Studierenden direkt bei der Arbeit beobachtet haben. D. h. wir müssen uns nicht auf die Selbstausskunft verlassen, die im Kontext von Sprachunterricht durch den Faktor der sozialen Erwünschtheit stärker verzerrt sein könnte, da die Studierenden in der Regel wissen,

Abb. 5: Anteil der Typen von Ressourcen an Satzbearbeitungen.

dass die Dozentinnen und Dozenten gerne hören, dass sie keine automatischen Übersetzungsprogramme nutzen. In diesem Sinne können die hier erhobenen Daten als Ermutigung für die Wörterbucharbeit verstanden werden: Die Studierenden aus unserer Studie scheinen tatsächlich sehr häufig Wörterbücher zu nutzen.

Im folgenden Abschnitt soll nun der Frage nachgegangen werden, ob für diese unterschiedlichen Ressourcen – v. a. Wörterbücher vs. Suchmaschinen – verschiedene Suchstrategien eingesetzt wurden.

4.2 Suchstrategien

Wie bereits oben erwähnt, können wir in unseren Studiendaten mehr als 2200 Suchaktionen untersuchen. Im Fokus dieses Abschnittes stehen nun die Ergebnisse zu den Suchstrings und ihrem Verhältnis zu den Typen der aufgerufenen Ressourcen.

Von der Sprache her sind die meisten Suchstrings deutsch, gefolgt von Suchstrings in der Ortssprache (vgl. Abb. 6). Die Anwendung der Ortssprache ist insofern auffällig, da in dieser Studie keine Übersetzungsaufgabe gestellt wurde, sondern eine Verbesserung deutscher Sätze gefordert war. Dieses Hin und Her mit den Sprachen (vgl. Abb. 7) scheint aber damit zusammenhängen, dass die Studierenden ihre

Abb. 6: Suchstringsprachen.

Abb. 7: Konsultationsverlauf pro Satz (Sequenzplot): Suchstringsprachen.

Tab. 4: Beispiel Suchsequenz mit unterschiedlichen Sprachen (ProbandIn S-16, Satz: „Ich habe die Hose viel zu klein gekauft. Jetzt muss ich nochmals ins Geschäft zurück und sie **wechseln**.“).

Name Ressource	Suchstring
Pons Translator.de	Ich habe die Hose viel zu klein gekauft. Jetzt muss ich nochmals ins Geschäft zurück und sie wechseln
Pons Dictionary.de	wechseln
Pons Translator.de	Ich habe die Hose viel zu klein gekauft. Jetzt muss ich nochmals ins Geschäft zurück und sie zu wechseln
Pons Dictionary.de	cambiar
Pons Dictionary.de	umtauschen
LEO Dictionary	umtauschen
Pons Translator.de	Ich habe die Hose viel zu klein gekauft. Jetzt muss ich nochmals ins Geschäft zurück und sie umtauschen
Pons Translator.de	Ich habe die Hose viel zu klein gekauft. Jetzt muss ich nochmals ins Geschäft zurück und sie die Hose umtauschen

Muttersprache als Vergewisserungsinstanz heranziehen und/oder den Interferenzfehlern dadurch auf die Spur kommen wollen, dass sie den deutschen Ausgangssatz in ihre Muttersprache rückübersetzen und dann bilinguale Ressourcen verwenden. Diese Strategie funktioniert in einigen Fällen auch sehr gut, wie das Beispiel in Tabelle 4 zeigt. Interessanterweise gehen Probandinnen und Probanden in Braga zusätzlich manchmal über Englisch als Drittsprache. In den Filmen kann man sehen, dass sie das v. a. dann machen, wenn sie erkennen, dass die konsultierten bilingualen Ressourcen deutsch-portugiesisch nicht gut funktionieren (z. B. in einem automatischen Übersetzungsprogramm), aber eine Übersetzung von deutsch-englisch in einem ersten Schritt und englisch-portugiesisch vielversprechender ist.

Unterschiedliche Typen von Ressourcen sind in der Regel für unterschiedliche Arten von Suchanfragen konzipiert. So sollte man beispielsweise in den meisten Wörterbüchern keine ganzen Sätze eingeben, wohingegen automatische Übersetzungsprogramme umso besser funktionieren, je mehr Kontext gegeben ist. Die Frage ist aber, ob dies den Studierenden auch bewusst ist. Wir möchten im Folgenden also der Frage nachgehen, ob wir diese Bewusstheit bei unseren Probandinnen und Probanden zeigen können.

Bei einer ersten Analyse fiel bereits auf, dass in Suchmaschinen oft der gesuchte sprachliche Ausdruck mit metasprachlichen Begriffen verbunden wurde, z. B. in folgenden Beispielen aus Rom (alles Suchstrings, die in „google.it“ eingegeben wurden):

- „wenn significato“
- „costruzione con wann“
- „einverstanden frasi“

- „handeln coniugazioni“
- „wenn costruzione“
- „beendet esempi“

Um nun überprüfen zu können, ob dies wirklich systematisch so ist, mussten die Suchstrings entsprechend typologisiert und annotiert werden. Zunächst einmal wurden die Suchstrings in drei Hauptklassen unterteilt:

- Unter die *metasprachlichen Begriffe* fallen Anfragen wie „Konjunktiv 2 mit wenn“
- Unter die *nicht-sprachlichen Anfragen* fallen die Namen von Ressourcen
- Unter die *sprachlichen Anfragen* fallen Anfragen nach konkreten sprachlichen Strukturen.

Die letzte Kategorie ist von Interesse, wenn wir überprüfen wollen, ob die Studierenden die Arten von sprachlichen Anfragen je nach Typen von Ressourcen unterschiedlich ausgestalten. Um dies prüfen zu können, haben wir die sprachlichen Anfragen in zwei Klassen eingeteilt:

- *einzelne Wortklassen* (mit entsprechenden Unterkategorien, die aber hier nicht berichtet werden) vs.
- *Komplexe* (auch mit entsprechenden Unterkategorien, die hier keine Rolle spielen).

Mit Komplexen sind dabei sowohl ganze Sätze als auch Phrasen etc. gemeint. Sinnvoll wäre es dabei, wenn die Studierenden bei den Wörterbüchern eher nach ein-

Abb. 8: Suchanfragen nach einzelnen Wortklassen vs. Komplexen im Pons Wörterbuch, Linguee, Pons Translator und Google Translator.

zelnen Wörtern suchen würden, d. h. nach einzelnen Wortklassen, wohingegen sie in automatischen Übersetzungsprogrammen eher nach umfangreicheren sprachlichen Strukturen, d. h. Komplexen, suchen sollten. Dass sie genau dies tun, zeigt sich sehr deutlich in unseren Daten (vgl. Abb. 8).

Auch hier scheinen sich die Studierenden besser auszukennen, als man es ihnen manchmal zuschreibt: Sie unterscheiden sehr klar nach den Funktionalitäten der einzelnen Tools und suchen wenig komplexe Phrasen im Wörterbuch und kaum einzelne Wörter im Übersetzungsprogramm.

4.3 Faktoren für die Korrektheit der Verbesserung

Interessant war für uns zu untersuchen, ob wir systematische Einflussfaktoren identifizieren können, die die Korrektheit der Verbesserungen beeinflussen. Ein Gedanke war beispielsweise, dass es einen positiven Einfluss auf die Satzverbesserung haben könnte, wenn viele Ressourcen verwendet werden. Dies zeigt sich allerdings nicht (vgl. Abb. 9). Sowohl bei korrekten Verbesserungen, bei falschen Korrekturen, bei Zweifelsfällen sowie bei gar nicht erfolgter Verbesserung ist der Mittelwert der Anzahl der konsultierten Ressourcen sehr ähnlich. Genauso hat auch die Bearbeitungsdauer pro Satz keinen Einfluss auf die Verbesserung. Auch die Position des Satzes in der Studie hat keinen Einfluss auf die Korrektheit. Es wurden also nicht eher die ersten Sätze korrekt gelöst und die letzten nicht mehr. Auch insgesamt können wir keinen sichtbaren Einfluss unserer Zeitwarnung (dass

Abb. 9: Anzahl verwendeter Ressourcen aufgeteilt nach korrekter Verbesserung, Zweifelsfall, falsch korrigiert oder nicht korrigiert (kein Eintrag).

Abb. 10: Wörterbuchanteil an konsultierten Ressourcen und Anteil korrekt verbesserter Sätze.

nur noch 15 Minuten zur Verfügung stehen) messen. Vielmehr scheint es so gewesen zu sein, dass es einfach zu verbessernde Sätze gab, die auch mit wenig Suchen und in kurzer Zeit korrekt zu verbessern waren, dagegen aber andere, die auch mit langer Gesamtzeit und vielen aufgerufenen Ressourcen für die Studierenden nicht leicht zu korrigieren waren.

Einen deutlichen Einfluss auf die Korrektheitsrate hat der Typ der verwendeten Ressourcen. Dabei sind insbesondere zwei Dinge auffällig. Erstens sind jene Teilnehmenden erfolgreicher, die mehr Wörterbuchressourcen benutzen. Der Zusammenhang ist in Abbildung 10 grafisch dargestellt. In Abbildung 10 lässt sich außerdem ablesen, dass dieser Zusammenhang insbesondere von den Teilnehmenden aus Braga gespeist wird, die nur wenig Sätze korrekt überarbeiten konnten. Eine weitere Unterteilung der Wörterbuchressourcen zeigt, dass Wörterbücher mit Paralleltexten tendenziell zu noch besseren Ergebnissen führen. Diesem Zusammenhang müssen wir allerdings noch genauer nachgehen, bevor sich verlässliche Aussagen treffen lassen. Zweitens lässt sich zeigen, dass die Probandinnen und Probanden, die mehr auf den Einsatz automatischer Übersetzungsprogramme setzten, schlechtere Überarbeitungsergebnisse erzielt haben. Dies sind gute Nachrichten für die Lexikografie: Im Kontext unserer Studie zeigt sich recht deutlich, dass die Verwendung von Wörterbuchressourcen messbar erfolgsversprechender ist – insbesondere im Gegensatz zu automatischen Übersetzungsprogrammen.

Bei der qualitativen Durchsicht der Bildschirmfilme und TAPs schien es uns außerdem sehr auffällig, dass manche Studierende sehr schnell zwischen unterschiedlichen Ressourcen wechseln und dabei den Eindruck vermitteln, gar nicht

Tab. 5: Ausschnitt aus dem ins Deutsche übersetzten TAP von R-07 zum Satz „Obwohl sich der Junge beeilt hat, hat er die U-Bahn **verloren**.“ (Hervorhebungen der AutorInnen).

also ich lese den ersten Satz vor [sie liest den Satz laut vor] (Obwohl sich der Junge beeilt hat, hat er die U-Bahn verloren) ahm das fettgedruckte Wort ist verloren also das Verb ahm es gibt zwei Sätze der eine ist der Hauptsatz und der andere ist der Nebensatz ahm als allererste suche ich das Verb verloren ich bevorzuge Google Chrome # ahm # normalerweise verwende ich PONS als # also Onlinewörterbuch weil es auch den Kontext und den Gebrauch angibt und auch einige Beispielsätze also ich stelle es auf Deutsch Italienisch ein okay

als verloren ## das bedeutet # okay also es ist ahm ein Partizip Perfekt des Verbs verlieren also perdere also Partizip Perfekt perduto perso also # ho perso il treno also nonostante il ragazzo (hat beeilt) ich suche es weil ich mir nicht sicher bin was es bedeutet also # beeilt (affrettarsi) okay (sich beeilen) also nonostante il ragazzo si sia affrettato ha perso il treno # *vielleicht liegt der Fehler bei der Tatsache dass verloren in anderen Kontexten gebraucht wird also ich suche andere Gebrauchskontexte oder ein Synonym des Verbs verlieren* also # ich suche verlieren und hier steht (perdere perdere la testa # disperdersi) okay

Ich suche vom Italienischen ein Synonym des Verbs perdere ins Deutsche also hier italienisch Deutsch ich suche perdere okay also (verlieren verlegen smarrire) # ahm für perdere il treno steht hier (verpassen) also ich kann verpassen anstelle von verlieren benutzen und ahm okay hier steht (Zug) also okay also ich verwende verpassen zur *Sicherheit suche ich das Partizip Perfekt* also suche ich eine Internetseite die REVERSO heisst

das Zeta [?] ## also ich suche coniugazione verbi tedeschi auf Google und hier haben wir REVERSO ich suche eben verpassen # also (verpasst) okay ich schreibe den Satz rechts wieder also obwohl dich der Junge beeilt hat hat der die U-Bahn also verpasst [sie schreibt den Satz fertig] okay ich mache mit dem zweiten Satz weiter

richtig zu lesen und auch schnell aufzugeben, wenn sie nicht unmittelbar eine gute Idee zur Verbesserung finden. Wiederum andere scheinen sich sehr konstruktiv in die Fehlersuche zu vertiefen und versuchen immer wieder, das gleiche Problem von unterschiedlichen Seiten neu anzugehen. Zwei Beispiele aus Rom sollen diese unterschiedlichen Arten des Vorgehens auf Basis zweier Ausschnitte aus TAPs verdeutlichen: In Tabelle 5 zeigen wir ein Beispiel einer Studierenden, die sehr ausdauernd recherchiert und Schritt für Schritt der Lösung näherkommt. Wenn sie keine gute Lösung findet, sucht sie z. B. nach anderen Gebrauchskontexten, nach Synonymen etc. (in der Tabelle kursiviert). Sie scheint also sowohl sorgfältig zu lesen, als auch gute Strategien zu haben, das Problem einzugrenzen. Dagegen zeigt die Studierende in Tabelle 6 ein anderes Verhalten. Wenn sie nicht relativ schnell eine Lösung findet, geht sie zum nächsten Satz. Interessant ist dabei auch, dass sie vom Computer oft in der dritten Person spricht (in der Tabelle unterstrichen). Sie sagt also nicht: „Ich finde nichts“, sondern „es“ oder „es findet nichts“. Damit verlagert sie die Akteursrolle hin zur Maschine.

Diese Beispiele scheinen sehr unterschiedliche Muster zu verdeutlichen. Bei einer solchen genauen Inspektion von Einzelbeispielen ist allerdings die Gefahr groß, generelle Tendenzen aus ihnen abzulesen, die sich im Gesamtdatensatz viel-

Tab. 6: Ausschnitt aus dem ins Deutsche übersetzten TAP von R-02 zu verschiedenen Sätzen (Hervorhebungen der AutorInnen).

also muss ich sehen was es regiert ob es den Akkusativ oder den Dativ regiert oder das ahm Verb
es ist eine Interjektion also # ahm wahrscheinlich wenn es nicht ahm wenn es nicht Dativ ist ist es Akkusativ also ahm es wäre hmmm wahrscheinlich ist es so # ich schreib den Fehler ab
okay aber um sicher zu sein suche ich welchen Kasus die die Präposition mit regiert indem ich es auf einer anderen Webseite suche oder auf einer anderen Seite # okay
ich suche eine Webseite die mir erlaubt okay # ich suche ahm aber das geht nicht ahm
ich suche ### ahm mir fällt nichts <in> okay okay es regiert # regiert den ahm # nein es regiert den Dativ <i>ich weiss es nicht ich mach weiter</i>
also ich mach weiter> <i>ich mach mit dem nächsten Satz weiter [...]</i>
also hmmm ich mach aber ich hab den nächsten Satz erblickt # ich glaube ich habe den Fehler verstanden auch wenn ich eine Präposition nicht erinnere also suche ich im Wörterbuch # okay
okay ich versuche ahm ihn zu übersetzen okay <i>ich verstehe was er meint</i> ahm aber ich verstehe nicht so müsste ich die Ordnung des Verbs umstellen # okay also übersetze ich ich meine ich übersetze und schreibe fertig # okay okay ahm
mit ahm okay also ist es nicht ich verstehe den Fehler bei der Präposition nicht es ist nicht mit es findet eine andere Präposition wie <?> ich muss mir kurz die Bedeutung der Präposition anschauen <??> aber um schneller zu sein suche ich direkt das italienische Äquivalent
ich öffne das Wörterbuch und suche ahm okay okay ahm um eine weitere Bestätigung zu haben suche ich es im Kontext also schreibe ich den Satz auf Italienisch oder zumindest sehr ähnlich # okay und ich suche eine Antwort wenn es geht hmmm okay
ich versuche den Satz zu ignorieren weil ich ihn falsch geschrieben habe okay genau wie ich dachte die Präposition ist falsch also muss ich sie ersetzen #
<i>dann der nächste Satz # [...]</i>

leicht gar nicht bestätigen. D.h., man muss sicherstellen, dass Einzelbeispiele nicht überbewertet werden. Der Vorteil an den von uns erhobenen Daten ist, dass man von der qualitativen Inspektion einzelner Beispiele viele Analysen auch auf quantitativer Ebene verifizieren kann. Um nun systematisch schauen zu können, ob das sorgfältige Lesen tatsächlich einen messbaren Einfluss auf die Verbesserungsrate hat, haben wir analysiert, ob die mittlere Verweildauer auf Ressourcen einen Einfluss auf die Verbesserung hat. Tatsächlich ist dies der Fall: Bei korrekt gelösten Sätzen ist die zeitliche Verweildauer auf den Ressourcen höher als bei nicht korrekt verbesserten Sätzen (vgl. Abb. 11). Der Unterschied in Sekunden ist zwar nicht groß (ca. 16 zu ca. 18,5 Sekunden), allerdings muss beachtet werden, dass sich dieser Unterschied auf jeweils eine einzelne Ressourcenverwendung bezieht. Das bedeutet, dass sich diese kleinen Differenzen in den einzelnen Verwendungen im Laufe des Gesamtversuchs zu einem deutlich größeren Effekt ausweiten

Abb. 11: Boxplots für die Verweildauer auf Ressourcen bei korrekt vs. nicht korrekt verbesserten Sätzen.

können. Es scheint also so zu sein, dass sich allein über die Verweildauer auf den Ressourcen eine Art Reflexion oder Lösungskompetenz zeigt. Daneben zeigt sich natürlich auch der Wille, das Problem zu lösen bzw. hartnäckiger nachzuforschen, aber das allein würde ohne Kompetenz nicht zu einer erfolgreichen Verbesserung führen.

Anfangs wurde gesagt, dass das Sprachniveau an den drei Erhebungsorten sehr unterschiedlich war, zumindest scheinen das die unterschiedlichen Verbesserungsraten zu zeigen (vgl. Abb. 1). Interessanterweise wird dieser Unterschied fast gespiegelt in den unterschiedlichen Verweildauern auf einzelnen Ressourcen: In Braga verweilen die Studierenden kürzer auf den konsultierten Ressourcen als die Studierenden in Santiago de Compostela und Rom (Abb. 12). Die (Sprach-)kompetenz, die Sätze lösen zu können, scheint also auch direkt mit der Dauer zusammenzuhängen, die auf die Ressourcen verwendet wird. Ob dies nun zeigt, dass die Studierenden die angezeigten Inhalte nicht verstehen oder zu schnell aufgeben, müssen weitere qualitative Analysen zeigen.

Bei der qualitativen Sichtung der Filme entstand zusätzlich der Eindruck, dass zwar eine überwiegende Anzahl der Studierenden den ersten Treffer nach einer Google-Recherche anklickt (vgl. Abb. 4), dass es aber auch dabei deutliche Unterschiede gibt: Die einen suchen etwas, das Suchergebnis erscheint und sie klicken, ohne auf die anderen Ergebnisse zu schauen, direkt das erste Ergebnis an. Andere dagegen suchen, schauen sich mehrere Treffer an und wählen erst dann das erste Suchergebnis aus. Dies führt zwar im Ergebnis beides zur Auswahl des ersten Google-Treffers, es sind aber doch qualitativ zwei unterschiedliche Handlungen:

Abb. 12: Boxplots für die Verweildauer auf Ressourcen bei korrekt vs. nicht korrekt verbesserten Sätzen unterschieden nach Erhebungsorten.

Abb. 13: Mittlere Verweildauer auf der Google-Ergebnisseite bei korrekt vs. nicht korrekt verbesserten Sätzen.

Während im ersten Fall die Auswahl unreflektiert erfolgt, ist sie im zweiten eine eher qualifizierte Auswahl. Auch diesem Eindruck haben wir versucht, quantitativ auf die Spur zu kommen. Wir haben dafür analysiert, ob die Zeit, die auf der Google-Ergebnisseite verbracht wird, einen Einfluss auf die Korrektheitsrate hat. Tat-

sächlich lässt sich in den Daten sehen, dass die Teilnehmenden bei korrekt gelösten Sätzen während der Satzbearbeitung durchschnittlich länger den Google-Ergebnisseiten verweilen, als wenn sie den Satz nicht korrekt überarbeiten (siehe Abb. 13). Der Mittelwertsunterschied beträgt hier 3,4 Sekunden und ist damit größer als der Unterschied in der Verweildauer auf Ressourcen bei korrekt vs. nicht korrekt gelösten Sätzen (dort ca. 2,5 Sekunden, vgl. Abb. 11).

Zusammenfassend lässt sich also sagen, dass wir v. a. zwei systematische Einflussfaktoren bestimmen konnten, die die Korrektheit der Verbesserungen bestimmen. Zum einen den Typ der verwendeten Ressourcen: Bei der Verwendung von Wörterbüchern ist eine korrekte Verbesserung in unserer Studie wahrscheinlicher als bei der Verwendung von automatischen Übersetzungsprogrammen. Zum zweiten das sorgfältige Lesen: Bei korrekt verbesserten Sätzen wurde auf den aufgerufenen Ressourcen länger verweilt. Dies scheint ein Indikator für ein kompetenteres oder reflektiertes Vorgehen zu sein oder schlicht den Willen zu zeigen, eine Lösung finden zu wollen und zu können. Der Unterschied der Verweildauer bevor das erste Google-Ergebnis ausgewählt wird, ist zwar numerisch recht deutlich, statistisch aber nicht abzusichern.

Neben diesen Faktoren scheint es aber auch sehr wichtig zu sein, dass die Studierenden überhaupt auf der richtigen Spur zur Suche des Fehlers sind, also die richtige Hypothese haben. Dies soll hier nur kurz an einem abschließenden Beispiel gezeigt werden.

4.4 Auf der falschen oder richtigen Spur: Die Bedeutung der Ausgangshypothese

Einer der Stimulussätze in unserer Studie war: *„Er wohnt seit Jahren in Berlin und trotzdem **verliert** er **sich** immer noch.“* (Satz Nr. 22). Von den zehn Studierenden in Braga, die versuchten, diesen Ausgangssatz zu korrigieren, haben acht angenommen, dass das Verb „verlieren“ in diesem Kontext richtig ist und dass es darum geht zu klären, ob „verlieren“ mit oder ohne Reflexivpronomen verwendet wird oder ob die Wortstellung falsch ist. Alle weiteren Recherchen gehen von dieser Annahme aus. Ein Beispiel soll dies verdeutlichen: Der Proband B-08 gibt, nachdem er versucht hat, den Sinn des Satzes zu verstehen, „verliert sich“ in Google Translate ein. Er sagt dazu: *„er wohnt seit Jahren in Berlin und trotzdem verliert er sich dort (verliert er sich # sich verliert) # mmm weiss nicht ob es (sich verliert # verliert verliert sich) ist #“*. Nachdem nur „perdido“ als Äquivalent in Google Translate erscheint, erweitert er die Suche auf „verliert er sich immer noch“ mit dem Kommentar: *„er wohnt seit Jahren in Berlin und trotzdem verliert er immer noch) # glaube dass (sich) hier überflüssig ist #“*. Auch hier wird als Übersetzung nur „ele ainda perdeu“ ausgegeben. Der Proband erwartet aber wohl, dass „perde-se“ oder „perder-se“ erscheinen sollte. Nun ändert er die Sprachrichtung des Google Übersetzers und gibt als Suchstring „ele ainda se perde“ ein. Als Übersetzung erscheint daraufhin: *„er verliert noch“*. Darauf sagt er: *„exakt glaube dass es (sich) nicht braucht“*

und korrigiert den Ausgangssatz entsprechend (falsch): „*Er wohnt seit Jahren in Berlin und trotzdem verliert er immer noch.*“ Seine komplette Recherche ist also nur dieser einen Frage gewidmet – reflexiv oder nicht – und führt in diesem Fall zu absurden Recherchehandlungen.

In anderen Beispielen zeigt sich, dass eine falsche Ausgangshypothese sogar dazu führen kann, dass – obwohl die korrekte Lösung im Wörterbuch enthalten ist – sie nicht gesehen wird. Beim Ausgangssatz „*Ich möchte ein Stipendium beim DAAD bewerben.*“ (Satz Nr. 19). vermuteten wiederum einige, dass sich das Problem um den Status des Verbs (reflexiv oder nicht reflexiv) dreht. Dies ist auch richtig, dabei übersehen sie aber, dass auch die Präposition beachtet werden sollte. So wird z. B. bei Pons im Sprachenpaar Deutsch-Portugiesisch der Beispielsatz gezeigt: „*sie bewirbt sich bei der Stiftung um ein Stipendium*“. Diesen Eintrag ruft eine Probandin (B-02) auf und könnte damit alle Informationen sehen, die sie für eine erfolgreiche Korrektur braucht, sie konzentriert sich aber nur auf das Reflexivpronomen und sagt: „*ich schaue mal nach was es bedeutet (bewerben)# (candidatar-se a) # ok es ist ein Reflexives Verb# also # es fehlt (sich)*“. Und so korrigiert sie den Satz auch in „*Ich möchte mich ein Stipendium beim DAAD bewerben*“, obwohl die vollständige korrekte Lösung mit einem passenden Beispielsatz direkt in dem von ihr konsultierten Wörterbuch präsent war.

Die richtige Hypothesenbildung zu unterstützen, kann allerdings nur im Sprachunterricht selber geleistet werden. Denn wie dieses Beispiel demonstrieren kann: Wenn die Studierenden durch die falsche Ausgangshypothese etwas Falsches suchen oder die Fehler nicht in vollem Umfang erkennen, können selbst die passenden lexikografischen Inhalte wenig zu einer korrekten Lösung beitragen. Dies deckt sich mit den Erkenntnissen von Frankenberg-Garcia 2011: 121:

Although this is no doubt important, the present study leads me to believe that the starting point should be the language problem rather than the dictionary. If we want learners to use dictionaries well, it is important to begin by helping them become aware of language problems that they are not used to confronting. It is only when learners appreciate that there is more to using a foreign language than finding L1-L2 equivalents and spelling words correctly that they can be taught how to use dictionaries to their full potential.

5 Schluss

Was können wir nun aus den Ergebnissen unserer Studie lernen? Allgemein zeigt sich, dass unsere Probandinnen und Probanden eine große Bandbreite an Ressourcen benutzen, obwohl sie sich noch nicht auf einem fortgeschrittenen Lernerniveau befinden. Wörterbücher werden dabei deutlich häufiger benutzt als automatische Übersetzungsprogramme. Dabei unterscheiden die Studierenden sehr genau zwischen den einzelnen Typen von Ressourcen, verfügen also anscheinend über ausreichendes Basiswissen zur Differenzierung der unterschiedlichen Funktionalitäten.

Wir denken, dass der Sprachunterricht genau hier ansetzen sollte: Anstatt pauschal Annahmen darüber zu machen, welche Ressourcen heutzutage von Studierenden genutzt werden, könnten unsere Studiendaten erstens als Anlass genommen werden, mit eigenen Studierenden und Sprachlernenden darüber zu diskutieren, welche Ressourcen sie nutzen und wie ihre Strategien dabei sind. Zweitens sollte das zumindest laut dieser Studie vorhandene Basiswissen über unterschiedliche Typen von Sprachressourcen genutzt werden, um darauf aufbauend noch mehr Wörterbuchbenutzungskompetenz zu vermitteln. Die Vermittlung von Wissen über verschiedene Ressourcen, die Kenntnis darüber, wo man was nachschlagen kann oder sollte und konstruktive Hinweise zum Lesen von lexikografischen Informationen sollte die Studierenden aber immer dort abholen, wo sie stehen. Tut man dies nicht, wären solche pauschalen Hinweise vergleichbar mit Sprachunterricht, der das Ausgangsniveau der Lernenden unberücksichtigt lässt.

Empirische Untersuchungen wie die hier vorliegende Studie sind auch deshalb wichtig, weil viele Ergebnisse aus unserer Studie für die beteiligten Sprachlehrerinnen unerwartet waren: Das Verwenden der Landessprache, manchmal sogar Englisch als Drittsprache im Wechsel mit Deutsch, die weitgehend kompetente Unterscheidung von Wörterbüchern vs. Übersetzungsprogrammen, auf der anderen Seite aber auch der Einfluss des sorgfältigen Lesens und der richtigen Ausgangshypothese: All dies scheint im Nachhinein fast erwartbar, war es aber vorher nicht. Insofern scheint es uns sehr lohnend, zum einen die Ergebnisse der Studie in den Sprachunterricht einfließen zu lassen und die eigene Wörterbuchbenutzungskompetenz mit Studierenden zu diskutieren, zum anderen aber auch mehr Daten in ähnlicher Weise zu erheben, um zu untersuchen, ob sich diese Ergebnisse auch in anderen Ländern, bei anderen Sprachen und mit anderen Aufgaben bestätigen oder eben nicht.

Aus jeder Studie lernt man auch etwas für zukünftige. Was wir in Zukunft anders machen würden, sind vor allem zwei Dinge: Zum einen würden wir einen kurzen Sprachtest vor der Studie machen. Nur dann kann man wirklich erfahren, ob es einen klaren Zusammenhang zwischen Sprachkompetenz und Rechercheverhalten gibt. Wir vermuten dies zwar für unsere Probandinnen und Probanden in Braga im Kontrast zu denen in Santiago de Compostela und Rom, wirklich beweisen können wir es aber nicht, weil wir ihre Sprachkompetenz nicht direkt erhoben haben. Zweitens würden wir für zukünftige Studien eher eine Übersetzungsaufgabe stellen. Für diese Studie schien es uns sehr wichtig, die identische Aufgabe für alle drei Orte zu haben. Allerdings zeigt sich in den Daten vor allem durch das Hin- und Herspringen in den Sprachen, dass die Aufgabe relativ unnatürlich für die Studierenden war. Auch im Sprachunterricht sind solche Aufgaben auf diesem Lernerniveau eher ungewöhnlich. Um also eine noch authentischere Aufgabe zum Gegenstand zu machen, wäre eine Übersetzung besser. Der methodische Aufbau mit Screenrecording und Think-Aloud hat dagegen sehr gut funktioniert. Allerdings würden wir zukünftig das laute Denken vor Beginn des Tests, zumindest kurz, mit jedem Probanden bzw. jeder Probandin üben, um den Anfang zu erleichtern.

Mehr Studien, die in diese Richtung gehen, wären für den Sprachunterricht und auch für die Lexikografie in Zukunft aber sicher sinnvoll, denn wir denken wie auch Lynn Bowker: „the key [...] is for lexicographers to listen to users“ (Bowker 2012: 396).

6 Literatur

- Domínguez Vázquez, María José/Mirazo Balsa, Mónica/Vidal Pérez, Vanessa (2013): Wörterbuchbenutzung: Erwartungen und Bedürfnisse. Ergebnisse einer Umfrage bei Deutsch lernenden Hispanophonen. In: Domínguez Vázquez, María José (Hrsg.) *Trends in der deutsch-spanischen Lexikographie*, Frankfurt a. M. u. a.: Lang, 135–172.
- Domínguez Vázquez, María José/Valcárcel Riveiro, Carlos (2015): Hábitos de uso de los diccionarios entre los estudiantes universitarios europeos: ¿nuevas tendencias?“. In: Domínguez Vázquez, María José/Gómez Guinovart, Xavier/Valcárcel Riveiro, Carlos (Hrsg.): *Lexicografía de las lenguas románicas II. Aproximaciones a la lexicografía contemporánea y contrastiva*. Berlin/Boston: de Gruyter, 165–189.
- Frankenberg-García, Ana (2011): Beyond L1–L2 Equivalents: Where do Users of English as a Foreign Language Turn for Help? In: *International Journal of Lexicography* 24/1, 97–123.
- Granger, Sylviane (2012): Introduction: Electronic Lexicography – from challenge to opportunity. In: Granger, Sylviane/Paquot, Magali (Hrsg.): *Electronic Lexicography*. Oxford: Oxford University Press. S. 1–11.
- Hult, Ann-Kristin (2012): Old and New User Study Methods Combined – Linking Web Questionnaires with Log Files from the Swedish Lexin Dictionary. In: Torjusen, Julie Matilde/Fjeld, Ruth Vatvedt (Hrsg.): *Proceedings of the 15th EURALEX International Congress 2012*. Oslo, Norway. S. 922–928. [Unter: http://www.euralex.org/elx_proceedings/Euralex2012/pp922-928%20Hult.pdf; letzter Zugriff: 17. 04. 2018].
- Koplenig, Alexander/Meyer, Peter/Müller-Spitzer, Carolin (2014): Dictionary users do look up frequent words. A log file analysis. In: Müller-Spitzer, Carolin (Hrsg.): *Using Online Dictionaries*. Berlin/Boston: de Gruyter. 229–249.
- Koplenig, Alexander/Müller-Spitzer, Carolin (2014): General issues of online dictionary use. In: Müller-Spitzer, Carolin (Hrsg.): *Using Online Dictionaries*. Berlin/Boston: de Gruyter, 127–141.
- Levy, Mike/Steel, Caroline (2015): Language learner perspectives on the functionality and use of electronic language dictionaries. In: *ReCALL* 27/2, 177–196.
- Lew, Robert (2010): Users Take Shortcuts: Navigating Dictionary Entries. In: Dykstra, Anne/Schoonheim, Tanneke (Hrsg.): *Proceedings of the 14th Euralex International Congress*. Ljouwert: Afûk. 1121–1132.
- Lew, Robert (2011): Studies in Dictionary Use: Recent Developments. In: *International Journal of Lexicography* 24/1, 1–4.
- Lew, Robert (2015a): Opportunities and limitations of user studies. In: Tiberius, Carole/Müller-Spitzer, Carolin (Hg.): *Research into dictionary use / Wörterbuchbenutzungsforschung*. 5. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“. Mannheim: Institut für Deutsche Sprache. 6–16. [Unter: https://ids-pub.bsz-bw.de/files/3772/Research_into_dictionary_use-2015.pdf; letzter Zugriff: 17. 04. 2018].
- Lew, Robert (2015b): Research into the use of online dictionaries. In: *International Journal of Lexicography* 28/2, S. 232–253.
- Lew, Robert/Grzelak, Marcin/Leszkowicz, Mateusz (2013): How Dictionary Users Choose Senses in Bilingual Dictionary Entries: An Eye-Tracking Study. In: *Lexikos* 23, 228–254.
- Müller-Spitzer, Carolin (2014) (Hrsg.): *Using Online Dictionaries*. Berlin/Boston: de Gruyter.

- Müller-Spitzer, Carolin/Koplenig, Alexander (2014): Online dictionaries: expectations and demands. In: Müller-Spitzer, Carolin (Hrsg.): *Using Online Dictionaries*. Berlin/Boston: de Gruyter, 143–188.
- Nied Curcio, Martina (2013): Der Gebrauch zweisprachiger Wörterbücher aus der Sicht italienischer Germanistikstudierender. In: Gouws, Rufus H. et al. (Hrsg.): *Lexicographica. International annual for lexicography* 29, Berlin/Boston: de Gruyter, 129–145.
- Töpel, Antje (2014): Review of research into the use of electronic dictionaries. In: Müller-Spitzer, Carolin (Hrsg.): *Using Online Dictionaries*. Berlin/Boston: de Gruyter, 13–54.
- Welker, Herbert Andreas (2014): Empirical research into dictionary use since 1990. In: Gouws, Rufus H. et al. (Hrsg.): *Dictionaries. An international encyclopedia of lexicography. Supplementary volume: Recent developments with focus on electronic and computational lexicography*. Berlin/New York: de Gruyter, 531–540.
- Wolfer, Sascha et al. (2015): The effectiveness of lexicographic tools for optimising written L1-texts. In: *International Journal of Lexicography* 31/1 (2018), 1–28 [online first: 2015].