

Sascha Wolfer*, Frank Michaelis und Carolin Müller-Spitzer

Nutzungsstatistiken von linguistischen Online-Ressourcen: Potentiale und Limitationen

Usage statistics of linguistic online resources: potentials and limitations

<https://doi.org/10.1515/zfal-2021-2070>

Abstract: Dictionary usage research views dictionaries primarily as tools for solving linguistic problems. A large proportion of dictionary use now takes place online and can thus be easily monitored using tracking technologies. Using the data gathered through tracking usage data, we hope to optimize user experiences of dictionaries and other linguistic resources. Usage statistics are also used for external evaluation of linguistic resources. In this paper, we pursue the following three questions from a quantitative perspective: (1) What new insights can we gain from collecting and analysing usage data? (2) What limitations of the data and/or the collection process do we need to be aware of? (3) How can these insights and limitations inform the development and evaluation of linguistic resources?

Keywords: Usage research, web analytics, evaluation, dictionaries, online resources

1 Einleitung

In der Wörterbuchbenutzungsforschung interessieren wir uns dafür, wie Benutzer*innen Wörterbücher verwenden, was sie nachschlagen (Müller-Spitzer, Wol-

***Kontaktperson: Sascha Wolfer**, Leibniz-Institut für Deutsche Sprache, Lexik, Mannheim,
E-Mail: wolfer@ids-mannheim.de

Frank Michaelis, Leibniz-Institut für Deutsche Sprache, Lexik, Mannheim,
E-Mail: michaelis@ids-mannheim.de

Carolin Müller-Spitzer, Leibniz-Institut für Deutsche Sprache, Lexik, Mannheim,
E-Mail: mueller-spitzer@ids-mannheim.de

fer & Koplenig 2015), ob sie schnell die gesuchte Information finden (Müller-Spitzer, Michaelis & Koplenig 2014) oder welche Eigenschaften sie an Wörterbüchern besonders wichtig finden (Kosem et al. 2018). Wörterbücher werden daher primär als Werkzeuge gesehen – und zwar nicht erst seit das Verbreitungsmedium für die enthaltenen Informationen nicht mehr nur das Buch aus Papier ist (Verlinde, Leroyer & Binon 2010). Wie alle Werkzeuge sind sie dafür gedacht in einer konkreten Situation ein spezifisches Problem einer bestimmten Person zu lösen. Seit ein großer Teil von Benutzungshandlungen in digitaler Form stattfindet (vgl. Kosem et al. 2018: 102), wächst daher die Hoffnung, den Nutzer*innen von linguistischen Ressourcen viel umfassender „auf die Finger schauen“ zu können als das bspw. bei gedruckten Wörterbüchern möglich ist, deren Benutzung üblicherweise nicht von außen einsehbar stattfindet. Die Möglichkeiten, die sich mit modernen Trackingtechnologien bieten, scheinen geradezu endlos und könnten – so die Hoffnung – die Tür weit aufstoßen zu einer extensiven Optimierung der Benutzung von Wörterbüchern, Informationssystemen und anderen Formen der Wissenschaftsvermittlung. De Schryver (2003: 188) drückt diese Hoffnung folgendermaßen aus: „[Y]et the electronic environment makes possible the truly exciting, namely the potential to retrieve feedback without the user being aware of this“.

Die Daten, die bei der Nutzung von Online-Ressourcen durch Tracking-Technologien anfallen, sind selbstverständlich nicht nur für die akademische Lexikographie von Interesse. Auf kommerzieller Seite stehen meist monetäre Aspekte im Vordergrund. Hier sollen zum Beispiel Antworten auf die Fragen gefunden werden, welche Werbeanzeigen an welchem Ort, zu welchem Zeitpunkt und welcher Person präsentiert werden müssen, um möglichst viele Klicks auf diese Anzeigen (und damit Gewinn) zu generieren. Im wissenschaftlichen Bereich wird es zudem immer üblicher, Zugriffszahlen und ähnliche Statistiken für externe Evaluationen anzugeben.

1.1. Fragestellungen und Perspektive

Ausgehend von diesen Erwartungen und Wünschen lauten die zentralen Fragestellungen, denen wir in diesem Papier nachgehen wollen:

- Welche neuen Erkenntnisse können wir aus Nutzungsdaten zu linguistischen Online-Ressourcen gewinnen, die uns dabei helfen können, diese Ressourcen benutzerfreundlicher, d. h. zu besseren Werkzeugen zum Lösen sprachlicher Problemstellungen zu machen?
- Welche Limitationen der anfallenden Datenspuren müssen wir uns bei der Interpretation von so erhobenen Daten bewusst sein?

- Was bedeutet das einerseits aus Sicht derer, die für den Betrieb und die (Weiter)Entwicklung von Ressourcen zuständig sind, und andererseits für externe Perspektiven auf diese Nutzungsdaten (z. B. bei Evaluationen)?

Wir werden bei der Bearbeitung dieser Fragestellungen durchweg eine quantitativ-statistische Perspektive einnehmen. Das heißt, dass wir nur in wenigen Einzelfällen auf die Datenspuren individueller Nutzer*innen eingehen werden. Der Fokus liegt somit auf der Herausarbeitung von größeren Daten- und Effektmustern und weniger auf Einzelfallstudien. Diese Perspektive sehen wir u. a. deshalb als für unsere Fragestellungen geeigneter an, weil wir Aussagen über ganze Klassen von Individuen treffen wollen und mehr an generalisierten Mustern als an wenigen Einzelfällen interessiert sind. Wir gehen außerdem beispielhaft von einer bestimmten Ressource aus, an der wir die Fragen exemplifizieren werden: dem Online-Wortschatz-Informationssystem Deutsch (OWID) des Leibniz-Instituts für Deutsche Sprache (www.owid.de). OWID ist ein Wörterbuchportal, an dessen Entwicklung wir Autor*innen beteiligt sind. Wir haben in den letzten Jahren die Potentiale und Limitationen von Nutzungsstatistiken aus zwei Blickwinkeln verfolgt, die beide in den vorliegenden Artikel einfließen: Zum einen als Forschende in der empirischen (ressourcenübergreifenden) Wörterbuchbenutzungsforschung (s. bspw. Müller-Spitzer 2014; Wolfer et al. 2016), zum anderen als Entwickler*innen einer bestimmten linguistischen Online-Ressource, die wir kontinuierlich verbessern wollen und die als Angebot eines Leibniz-Instituts regelmäßig extern evaluiert wird. Je nach Entwicklungsmethodologie werden die Benutzer*innen mal früher, meistens eher später in den Entwicklungsprozess eingebunden, wenn überhaupt. Da wir – wie oben erwähnt – Wörterbücher primär als Hilfsmittel für unsere Benutzer*innen ansehen, sollten sie auch Teil des Entwicklungsprozesses sein.

1.2. Gliederung

Wir werden in Abschnitt 2.1 zunächst einige technische Grundlagen für das Sammeln von Nutzungsdaten darlegen. Wir werden auch versuchen abzuschätzen, wie umfänglich der anfallende Datenbestand die tatsächliche Nutzung der Online-Ressource abdeckt (Abschnitt 2.2). In Abschnitt 3.1 werden wir darauf eingehen, unter welchen Fragestellungen diese Art der Nutzungsbeobachtung bereits für linguistische Online-Ressourcen eingesetzt wurde. Im folgenden Abschnitt 3.2 werden wir einige Auswertungen zum Wörterbuchportal OWID vorstellen. Der Diskussionsabschnitt 3.3 dient dazu, die Potentiale und Limitationen ausgehend von den Analysen zu OWID zusammenzufassen. In unserem Fazit (Abschnitt 4)

leiten wir ab, welche alternativen Methodologien wir als geeignet ansehen, die Limitationen anzugehen, um über bloße Hypothesenbildung hinaus klar definierte Fragestellungen im Bereich der Benutzungsforschung untersuchen zu können. Zudem werden wir auch auf den Gegensatz zwischen internen und externen Perspektiven auf Nutzungsdaten in der Wissenschaft eingehen.

2 Grundlagen

Der Grundgedanke bei der Erhebung von Nutzungsstatistiken ist es immer, ein möglichst vollumfängliches Bild der Interaktion von Nutzer*innen mit einer Webseite zu erhalten. Wir wollen in diesem Abschnitt darstellen, welche Plattformen dafür genutzt werden und wie „vollumfänglich“ die Daten sind, die man auf diesem Weg gewinnen kann. Wir werden deshalb kurz auf die technischen Grundlagen von Server-Logging einerseits und Webanalyse-Plattformen andererseits eingehen. Wir werden dann Kennwerte dieser beiden Quellen vergleichen, um so einen Einblick in die potentielle Dunkelziffer bei der Nutzungserhebung anhand moderner Webanalyse-Plattformen zu gewinnen.

2.1 Nutzungsmessung

Jeder Webserver erstellt Aufzeichnungen davon, wie oft die auf ihm gespeicherten Ressourcen angefragt werden (sog. Requests). Ziel dieser Aufzeichnungen ist es weniger Nutzungsstatistiken zu erstellen, sondern eher Informationen über die Auslastung und den technisch einwandfreien Betrieb zu erhalten, z.B. ob es zu bestimmten Zeiten besondere Lastspitzen mit besonders vielen Requests pro Sekunde gab und ob der Server hinreichend für diese Last dimensioniert gewesen ist. Wie genau diese Requests geloggt werden, unterscheidet sich je nach Serversoftware und -konfiguration. In der Regel werden aber die Ortszeit des Servers sowie die URL der aufgerufenen Ressource geloggt. Die Anzahl der Requests kann dabei nicht ohne weiteres mit Seitenaufrufen gleichgesetzt werden, denn Webseiten bestehen nicht nur aus der eigentlichen HTML-Seite. Für die vollständige Darstellung im Browser werden weitere Ressourcen vom Server heruntergeladen, etwa CSS-Stylesheets, Bilder für Hintergründe, Logos oder JavaScript-Code. Jeder Abruf einer dieser Ressourcen führt zu zusätzlichen Requests. Ein einzelner Aufruf von www.owid.de führt auf dem Webserver zu etwa 20 Requests. Zum Vergleich: Ein Aufruf von www.duden.de führt zu etwa 110 Requests, einer von www.spiegel.de je nach Tagesinhalt zu bis zu 100. Die Zahlen können zudem je nach Browser-Caching abweichen.

Zur eher inhaltlichen Frage der Interaktion von menschlichen Benutzer*innen mit Webseiten sind Webanalyse-Plattformen¹ entwickelt worden. Während Server-Logs originär eher auf die technische Seite der Nutzungsmessung ausgerichtet sind, waren Webanalyse-Plattformen von Beginn an darauf spezialisiert, Benutzungshandlungen über mehrere Seiten eines Angebots hinweg oder mithilfe Dritter (z. B. *Google Analytics*) gar über mehrere Webangebote übergreifend aufzuzeichnen und zu untersuchen. Das Verhalten von Benutzer*innen wird mithilfe dieser Plattformen auf einer abstrakteren Ebene als den Server-Requests analysiert und soll den Betreiber*innen der Seiten bei Inhalts- und Designentscheidungen helfen. Typische Bestandteile solcher Webanalyse-Software sind Echtzeitanalysen von Benutzer*innen (welche Seite sie gerade aufgerufen haben), Übersichten über die Herkunft von Benutzer*innen sowie die für den Zugriff verwendete Technologie (Browser, Auflösung, Betriebssystem usw.). Auch eher auf E-Commerce ausgerichtete Maßzahlen wie die *Conversion* sind verfügbar, also der Anteil von Besucher*innen, die sich für einen Kauf in einem Online-Shop entscheiden. Die verbreitetste Webanalyse-Plattform ist *Google Analytics*. Diese Software setzen wir aus Datenschutzgründen nicht zur Untersuchung unseres Angebots ein, da von *Google Analytics* erhobene Nutzungsdaten immer auf Servern mit unbekanntem Standort gespeichert werden. Grundsätzlich ist es daher möglich, dass alle Daten, die über *Google Analytics* gesammelt werden, vom Mutterkonzern genutzt werden, um umfassende und Websites überspannende Tracking-Daten zu erheben.

Matomo, ehemals *Piwik*, ist dagegen eine Open-Source-Webanalytik-Plattform und eine Alternative zu *Google Analytics*. Auf Seiten mit der Domain *.de* ist *Matomo* hinter *Google Analytics* (allerdings mit einigem Abstand) das am häufigsten eingesetzte Werkzeug zur Erhebung und Auswertung von Website-Zugriffen². *Matomo* kann auf eigenen Servern installiert werden, was u. a. bedeutet, dass die erhobenen Daten ebenfalls auf diesen Servern abgelegt werden und nicht für Dritte einsehbar sind. Auch für *Matomo* gilt, dass der Funktionsumfang hauptsächlich auf E-Commerce-Websites ausgerichtet ist, was nicht bedeutet, dass es nur für diese Zwecke eingesetzt werden kann. *Matomo* stellt – wie auch *Google*

1 Den Terminus „Plattform“ haben wir von der Eigenbezeichnung von *Matomo* übernommen („web analytics platform“, <https://matomo.org> [letzter Zugriff am 1.4.2021]). Bei den von Gillespie (2010) aufgeführten Definitionskategorien, die er zum Terminus „platform“ aus den im *Oxford English Dictionary* (Stand April 2009) angegebenen Lesarten ableitet, passen auf *Matomo* u. E. allerdings am ehesten die komputationale und figurative. Wir danken einer*m anonymen Gutachter*in für den Hinweis, dass der Plattform-Begriff hier nicht ganz unproblematisch ist.

2 https://w3techs.com/technologies/segmentation/tld-de-/traffic_analysis [letzter Zugriff am 12.3.2021]

Analytics – ein sog. *Dashboard* zur Verfügung, auf dem über eine grafische Nutzeroberfläche auf bestimmte Auswertungen zugegriffen werden kann. Alle Analysen, die wir in diesem Beitrag präsentieren, basieren jedoch auf exportierten Rohdaten in Tabellenform. Bei der Auswertung verzichten wir somit auf die vorgefertigten Analysemöglichkeiten von Matomo. Dafür können wir eigene Filterungen und Analysen – wir verwenden hierzu die Software R (R Core Team 2021) – an unsere jeweiligen Fragestellungen detaillierter anpassen, als das über das Webinterface möglich ist. Für dieses Vorgehen haben wir uns auch deshalb entschieden, weil Webanalyse-Plattformen zwar für wissenschaftliche Online-Ressourcen eingesetzt werden können, der Fokus aber eher auf Anwendungen aus dem Bereich E-Commerce liegt – also Webangeboten, die auf ein technisch messbares Ziel, bspw. die Kaufentscheidung, ausgerichtet sind. Im Falle eines Wörterbuchportals wie OWID kann das Ziel nicht so eindeutig festgelegt werden. Das wiederum bedeutet, dass wir die Datenaufbereitung und -auswertung flexibler gestalten müssen als das mit den vorgefertigten Funktionen möglich ist. Diese Flexibilität ist u.E. am ehesten gegeben, wenn wir von den Rohdaten ausgehen.

Speziell für Matomo müssen wir im Folgenden eine bestimmte Terminologie zugrunde legen, die auch in den eben erwähnten Rohdaten verwendet wird. Wir benutzen hier die englischen Begriffe, um deren technische Dimension zu verdeutlichen. Ein *Visitor* ist eine die Webseite besuchende Person, allerdings nur insofern Matomo sie mittels Tracking-Techniken, wie z.B. Cookies, identifizieren kann. Das heißt, ein Visitor ist eigentlich ein Browser auf einem bestimmten Computer. Inwieweit dies mit einem konkreten Menschen zusammenfällt, können wir bei der Betrachtung der Daten nicht wissen. Jedem Visitor ist mindestens ein *Visit* zugeordnet, also eine zusammenhängende Sequenz von Aktionen, die der Visitor innerhalb des Webangebots durchgeführt hat. Die kleinste Einheit eines Visits ist die *Action*, also eine konkrete Aktion, die der Visitor in einem Visit getätigt hat. Bei OWID handelt es sich hier immer um Seitenaufrufe³. Einem Visitor können mehrere Visits zugeordnet sein. Jedem Visit können mehrere Actions zugeordnet sein.

Ein konkretes Beispiel zur Verdeutlichung: Visitor „5899296f781ad7ec“ (dieser Code wird zufällig und eindeutig vom System für einen bestimmten Browser vergeben) aus Medellin in Kolumbien griff am 13. März 2018 um 5:18 Uhr (lokale Zeit beim Benutzer: 23:18 Uhr) auf die Hauptseite von OWID zu (Action 1). Nach 12 Sekunden griff der Visitor auf die Ergebnisseite zur Suche „quälen“ zu und ver-

³ Bei anderen Webseiten wie bspw. E-Commerce-Angeboten könnte eine Action auch ein Kauf sein.

brachte dort 23 Sekunden (Action 2). Dann rief sie*er die *ellexiko*-Seite des Eintrags „Quälen“ auf (Action 3), verbrachte dort lediglich 2 Sekunden und besuchte dann die *ellexiko*-Seite des Eintrags „quälen“ (Action 4). Diese Action schließt Visit 1 dieses Visitors ab. Am 31. März 2018 besucht derselbe Visitor OWID erneut und wird von Matomo bei 3 Actions geloggt (Hauptseite, Suchergebnisseite zu „Knie“, *ellexiko*-Seite zu „Knie“). In den Matomo-Logs aus dem Jahr 2018 sind also 7 Actions von diesem Visitor verzeichnet, verteilt auf 2 Visits.

Im obigen Beispiel haben wir für Action 4 des ersten Visits keine Dauer angegeben. Das hat einen einfachen Grund: Die letzte Action eines Visits ist niemals mit einer Dauer versehen.⁴ Somit können wir Verweildauern nur für Visits berechnen, die mehr als eine Aktion umfassen, denn wenn ein Visit nur eine Action umfasst, ist diese Action gleichzeitig die letzte.

2.2 Dunkelziffer

Wir haben anfangs gesagt, dass wir erläutern werden, wie umfänglich die Daten zu Benutzerinteraktionen sind, die wir über Webanalyseplattformen ermitteln können. Dafür vergleichen wir nun die Zahlen von Server-Requests mit den von Matomo ermittelten Zugriffszahlen. Die beiden Herangehensweisen operieren auf technischer Ebene zwar unterschiedlich, aber sie ermöglichen uns einen Einblick in die Vollständigkeit der Matomo-Zahlen. Ein solcher Vergleich scheint uns auch deshalb interessant, da Benutzer*innen die Möglichkeit haben, eine „Do-Not-Track“-Option in ihren Browsereinstellungen zu aktivieren, die zumindest bei der am IDS installierten Matomo-Instanz dazu führt, dass das Tracking-Tool die Aktionen dieser Benutzer*in nicht verfolgen wird. Eine Auswertung auf der Basis von Server-Requests kann hingegen nicht von den Browsern der Besucher*innen unterbunden werden, da sie lediglich serverseitig geschieht. In einem gewissen Rahmen können wir somit eine „Dunkelziffer“ abschätzen, die erfasst, wie hoch der Anteil von Visitors ist, die vom eingesetzten Webanalyse-Tool nicht getrackt werden. Da die Unterschiede der beiden Zählweisen unseres Wissens selten dokumentiert werden, diese aber relevant für die Interpretation von Nutzungsstatistiken sind, zeigen wir im Folgenden die Abweichungen innerhalb von OWID. Um

⁴ Dass der letzten Action niemals eine Dauer zugeordnet ist, hat technische Gründe. Eine Dauer kann nur dann zugeordnet werden, wenn Action n durch die folgende Action $n + 1$ beendet wird. Verlässt ein*e Benutzer*in aber die Website (bspw. durch Schließen des Browserfensters oder Eingeben einer anderen URL), löst das keine Action aus, womit der vorhergehenden Action auch keine Zeit zugeordnet werden kann.

die beiden Operationalisierungen der Zugriffszahlen (Requests vs. Actions) vergleichen zu können, haben wir die Server-Requests so eingeschränkt, dass nur ein Request pro Seitenaufruf gezählt wird.⁵ Alle Abweichungen, die sich nun noch ergeben, sind darauf zurückzuführen, dass (1) entweder Besucher*innen das Tracking unterbunden haben oder (2) die Zugriffe von automatisierten Programmen (Webcrawler, Bots o.ä.⁶) stammen, die in großem Stil die Inhalte von Webangeboten abgreifen und nicht von Matomo erfasst werden.

Als Vergleichszeitraum entschieden wir uns für das komplette Jahr 2018. Weiterhin unterscheiden wir für diese Vergleiche drei Typen von Seiten auf OWID, die uns besonders geeignet schienen, da sie fest umrissene Funktionalitäten betreffen. Wir ermitteln die Dunkelziffer also nicht anhand aller URLs in OWID, sondern schätzen diese über die Teilmenge der folgenden drei interpretierbaren Kategorien ab:

- **Artikel**, also alle URLs, die die Zeichenfolge „/artikel/<Identifikationsnummer>“ enthalten.
- **Direktzugriff** ist definiert durch alle URLs, die das Muster „/nav/gehezu/<Suchstring>“ enthalten. Diese URL wird in Folge einer Eingabe in ein „Gehe zu Wort“-Eingabefeld aufgerufen.⁷
- **Hauptsuche**, also alle URLs, die die Zeichenfolge „/suche/wort?word=<Suchstring>“ enthalten. Diese Kategorie umfasst somit alle durchgeführten Hauptsuchen⁸.

⁵ Das ist notwendig, weil – wie wir zuvor ausgeführt haben – typischerweise mehr als ein Server-Request pro Seitenaufruf anfällt.

⁶ Grundsätzlich ist es möglich, Programme zu erstellen, die einen Browser und dessen menschliche Bedienung „nachahmen“. Solche Programme würden auch von Matomo getrackt. Allerdings stellen solche Programme im Feld der Applikationen, die automatisch Inhalte von Webseiten herunterladen, die Ausnahme dar.

⁷ Die URL <https://www.owid.de/nav/gehezu/Geduld+bringt+Rosen> führt bspw. direkt zum Artikel „Geduld bringt Rosen“ im Sprichwörterbuch [letzter Zugriff am 16.9.2020]. Das Eingabefeld „Gehe zu“ ist seit März 2020 nicht mehr Bestandteil der OWID-Oberfläche.

⁸ Die URL <https://www.owid.de/suche/wort?word=gehen> führt bspw. zu den Suchergebnissen zu „gehen“ [letzter Zugriff am 16.9.2020].

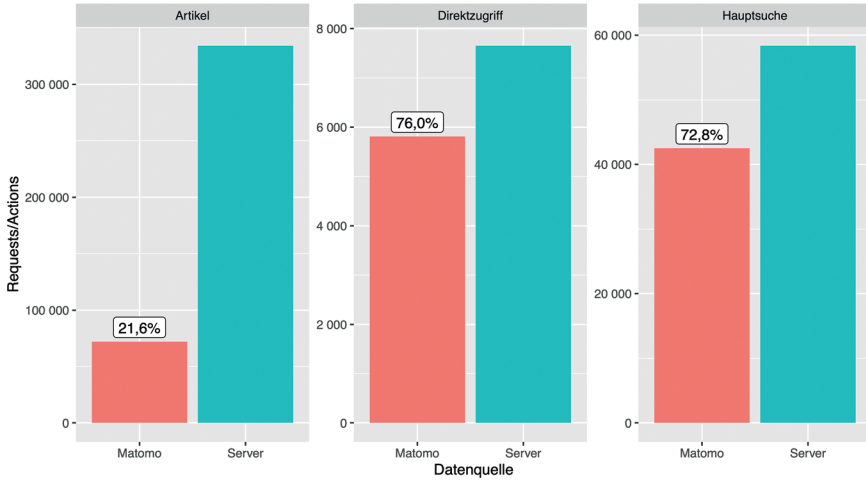


Abbildung 1: Vergleich von Matomo-Actions (linke Balken) und Server-Requests (rechte Balken) für die Kategorien Artikel, Direktzugriff und Hauptsuche. Auf der y-Achse ist die jeweilige Menge im Jahr 2018 abgetragen. Die Prozentwerte über den linken Balken geben das prozentuale Verhältnis der Mengen an.

Abbildung 1 zeigt den Mengenvergleich zwischen von Matomo erfassten Actions (linke Balken) und den unifizierten Server-Requests für die drei Kategorien Artikel, Direktzugriff und Hauptsuche. Die Prozentwerte über den linken Balken geben dabei jeweils an, wie viele Actions im Vergleich zu unifizierten Requests gezählt wurden. Die Dunkelziffer ist somit der Abstand zwischen der Höhe des linken Balkens und der Höhe des rechten Balkens (oder, in Prozenten ausgedrückt, 100 % minus den angegebenen Prozentwert).

Es fällt auf, dass sich für die Kategorien Direktzugriff und Hauptsuche – bei- des Zugriffsstrukturen und keine inhaltlichen Kategorien – der Wert in einem ähnlichen Bereich befindet. Dort wurden 24 % bzw. 27,2 % der Zugriffe nicht durch Matomo detektiert. Bei der inhaltlichen Kategorie Artikel ergibt sich fast ein umgekehrtes Bild. Dort wurden lediglich 21,6 % aller Zugriffe durch Matomo erfasst, die Dunkelziffer liegt hier bei 78,4 %. Wir haben keine definitive Antwort auf die Frage woher dieser beträchtliche Unterschied stammt, doch wenn wir uns nochmal vor Augen führen, woher Unterschiede zwischen Action- und Request-Anzahl stammen können, ergibt sich zumindest ein Verdacht. Wir gehen nicht davon aus, dass die unterschiedlichen Dunkelziffern von einem systematischen Unterschied beim Browser-seitigen Unterbinden des Trackings stammen. Wäre das der Fall, würden Visitors selektiv häufiger beim Aufrufen von Artikeln das Tracking unterbinden als beim Direktzugriff oder beim Suchen. Wahrscheinlicher

ist u.E. eine andere Ursache, nämlich das automatisierte „Abgreifen“ von inhaltlichen Ressourcen auf OWID durch automatisierte Programme. Diese Programme sind in der Regel so programmiert, dass der JavaScript-Code, der für das Matomo-Tracking notwendig ist, nicht ausgeführt wird. Ein automatisiertes Speichern von Inhalten ergibt aber nur auf Seiten Sinn, die auch inhaltlich ausgerichtet sind, und das sind unter den drei betrachteten Kategorien nur die Artikel und eben nicht die Hauptsuche oder der Direktzugriff. Für diesen Verdacht spricht ein weiteres Indiz, nämlich dass es Zeiträume gibt, in denen der Unterschied zwischen erfassten Actions und unifizierten Requests gerade bei den Artikeln besonders hoch ist. Im hier untersuchten Datensatz ist das der Mai 2018. Dort beträgt das Verhältnis von Actions vs. unifizierte Requests 14,6 % (im selben Zeitraum für Hauptsuche: 78,7 %; für Direktzugriff: 77,2 %) während es in allen anderen Monaten durchschnittlich bei 24,5 % liegt (Hauptsuche: 73,0 %; Direktzugriff: 75,4 %). Das heißt, dass im Mai 2018 die Dunkelziffer für die Artikel besonders hoch ist. Dies spricht u.E. für eine zeitlich konzentrierte Download-Aktion von OWID-Artikelinhalten.

Wie können wir nun insgesamt die Dunkelziffer beim Matomo-basierten Tracking abschätzen? Das breite Spektrum von 24 % bis 78 % können wir mit einiger Sicherheit eingrenzen: Wenn wir davon ausgehen, dass die hohe Dunkelziffer für die Artikel auf automatisierte Zugriffe zurückzuführen ist, ist die Dunkelziffer für menschliche Zugriffe eher bei 24 % anzusetzen. Wir würden also in etwa davon ausgehen, dass wir über die Matomo-Daten etwa 75 % aller tatsächlichen Zugriffe auf OWID erfassen. Dies müssen wir bei der Interpretation der später vorgestellten Ergebnisse berücksichtigen.

3 Nutzungsstatistiken für linguistische Ressourcen

3.1 Bisherige Forschung

Für Wissenschaftler*innen (nicht nur aus dem Bereich der Lexikographie) kann es aus mehreren Gründen interessant sein, wie sprachbezogene Online-Ressourcen benutzt werden. Wir nennen hier exemplarisch Forschungsarbeiten aus dem Bereich der digitalen Lexikographie. Eine ganze Reihe von Untersuchungen beschäftigt sich dabei mit den Zugriffszahlen auf einzelne Wörterbuchartikel. Dabei sollen allgemeine Aussagen getroffen werden, wie Ressourcen-externe Variablen (allen voran die Häufigkeit von Wörtern in allgemeinsprachlichen Korpora) zusammenhängen mit der Nachschlagefrequenz in Wörterbüchern. So konnten

bspw. de Schryver et al. (2006) in einem Swahili-Englisch-Wörterbuch keinen Zusammenhang zwischen Korpusfrequenz und Nachschlagehäufigkeit nachweisen. Jedoch konnten Müller-Spitzer et al. (2015) mit Auswertungen anhand von Log-Files des freien Wörterbuchs Wiktionary unter Verwendung einer anderen Methodik einen solchen Zusammenhang feststellen. In einer diese Methodik replizierenden Untersuchung von de Schryver et al. (2019) konnte der entsprechende Zusammenhang dann auch im ursprünglich untersuchten Swahili-Englisch-Wörterbuch gezeigt werden. Der Zusammenhang zwischen Korpusfrequenz und Nachschlagehäufigkeit in einem Wörterbuch kann insofern als stabil erachtet werden: Im Korpus häufige Wörter werden auch häufig nachgeschlagen. Dies hat u.a. Auswirkungen auf die Selektion von Wörtern in allgemeinsprachlichen Wörterbüchern: Wenn häufige Wörter auch häufig nachgeschlagen werden, ist es sinnvoll, diese Wörter (prioritär) zu beschreiben. Außerdem – so zeigten Müller-Spitzer et al. (2015) – werden polyseme Wörter unabhängig von der Korpusfrequenz häufiger nachgeschlagen als monoseme. Ein weiterer Effekt, der sich bei dieser Auswertung von Wiktionary-Logfiles zeigte, ist einer der temporären sozialen Relevanz: Wenn Wörter plötzlich eine prominente Rolle in gesellschaftlichen Diskussionen einnehmen (oder auch einfach in populären Quiz-Sendungen nach ihnen gefragt wird) hat das oft einen deutlichen, aber kurzen Einfluss darauf, wie häufig das Wort nachgeschlagen (siehe zu ähnlichen Effekten auch Wolfer 2020).

Andere Untersuchungen nehmen nicht die Nachschlagehäufigkeit von Stichwörtern in den Fokus, sondern Benutzungshandlungen. So werten Tiberius & Nierstadt (2015) in ihrer Arbeit die internen Log-Files des „Algemeen Nederlands Woordenboek“ (ANW, Dictionary of Contemporary Dutch) aus und stellen fest, dass Benutzer*innen, sofern sie überhaupt wörterbucheigene Suchen verwenden und nicht direkt von Google zu den Artikeln geleitet werden, erweiterte, komplexe Suchen seltener als erwartet verwendeten, was für die Konzeption eines Redesigns der Benutzeroberfläche des ANW herangezogen werden soll.

Basierten diese Untersuchungen entweder auf Server-Logs oder vorverarbeiteten Log-Files (im Falle des Wiktionarys), wurden auch Webanalyse-Plattformen dazu eingesetzt, das Verhalten von Nutzer*innen auf Online-Sprachressourcen zu untersuchen. Falke & Ripp (2018) interessierten sich bspw. insbesondere für das Suchverhalten und Selektionsverhalten von Nutzer*innen im grammatischen Informationssystem *grammis* und setzten dabei ebenfalls Matomo ein. Auch Google Analytics wurde für die Nutzungsanalyse von Online-Wörterbüchern eingesetzt. So untersuchen Lorentzen & Theilgaard (2012) die Auswirkungen verschiedener Google-Indizierungseinstellungen (*search engine optimization*) auf die Benutzerzugriffe von „ordnet.dk“ (Den Danske Ordbog, DDO) und die verwendeten Suchbegriffe. Um darüberhinausgehende Aussagen zum Nutzungsverhalten machen zu können, wurde deren Zugriffsanalyse durch einen Online-Fragebogen ergänzt.

Im Folgenden werden wir darstellen, wie wir Matomo-Benutzungsdaten einsetzen können, um mehr darüber herauszufinden wie verschiedene Funktionalitäten und Ressourcen innerhalb des Wörterbuchportals OWID benutzt werden. Dabei interessieren wir uns auch dafür, wie sich die Zugriffe über die Zeit entwickeln, wie lange Benutzungshandlungen typischerweise dauern und ob wir bestimmte Nutzungsmuster und -gruppen identifizieren können.

3.2 Zugriffstatistiken für OWID

Im Folgenden werden wir die mit Matomo gesammelten Daten für die Jahre 2018 und 2019 verwenden, aufzeigen, wie wir diese Daten aufbereitet haben, und zusammenfassen, welche Folgen diese Aufbereitungen haben.

3.2.1 Datensatz

Wir verwenden für die folgenden Analysen die Gesamtdaten für OWID aus den Jahren 2018 und 2019. Diese umfassen insgesamt 529 049 Actions in 96 828 Visits von 46 618 Visitors. Wir schließen vorab alle Zugriffe vom Leibniz-Institut für Deutsche Sprache aus, da wir uns für die Nutzung unserer Ressource außerhalb des Kollegiums interessieren. Das betrifft 36 024 Actions (6,8%), 3 925 Visits (4,1%) und 834 Visitors (1,8%). Der zu analysierende Datensatz umfasst somit 493 025 Actions, 92 903 Visits und 45 784 Visitors.

Analysen können auf unterschiedlichen „Ebenen“ des Datensatzes ansetzen. Die drei Ebenen sind in absteigender Nennung: Visitor, Visit, Action (zur Terminologie siehe Abschnitt 2.1). Alle Variablen, die auf der Ebene eines Visits definiert sind, wie beispielsweise Herkunft des Zugriffs (Nation, Stadt), oder der verwendete Browser usw. werden dabei an die unterste Ebene, die Actions, vererbt. Dieses Vererbungsprinzip gilt nicht zwischen Visitor und Visits: Zwar ist jeder Visit einem Visitor zugeordnet, aber nicht alle Variablen des Visitors werden an alle Visits vererbt. Das gilt z.B. für die Herkunft des Zugriffs: Bewegt sich das Gerät, mit dem ein Visitor auf OWID zugreift, verändert sich logischerweise auch der geloggte Ort. Ein konkretes Beispiel: der Visitor mit den meisten geloggten Städten im Datensatz hat über 96 Visits hinweg von 38 verschiedenen Städten⁹ aus auf OWID zugegriffen.

⁹ Die Angabe von 38 Städten basiert auf den Matomo-Daten. Diese kann unterschiedliche Ursachen haben: Es muss nicht zwingend der menschliche Benutzer sein, der sich in 38 verschiedenen

Um die vorhandene Datenmenge in den Griff zu bekommen und analytisch den Überblick zu behalten, kategorisieren wir die Actions im Datensatz basierend auf der assoziierten URL. So können wir die einzelnen Handlungen der Visitors bestimmten abstrakteren Aktionstypen unterordnen, die spezifisch für das Wörterbuchportal OWID sind. Wir unterscheiden dabei die folgenden Aktionstypen (die Typen sind absteigend nach ihrer Häufigkeit sortiert):

- **Artikel aufrufen:** Dieser Aktionstyp wird zugewiesen, wenn auf einen Wortartikel in einem der auf OWID verfügbaren Wörterbücher zugegriffen wurde (unabhängig davon, um welches der in OWID enthaltenen Wörterbücher es sich handelt). Auf diesen Aktionstyp entfallen 174 113 (35,3 %) aller Actions.
- **Suchen:** Wann immer eine Suchseite in OWID aufgerufen wurde, wird die Action diesem Aktionstyp zugeordnet, Seiten zu erweiterten Suchen sind hier ausgenommen. Wir bezeichnen diesen Typ im Folgenden als „Hauptsuche“. Auf diesen Aktionstyp entfallen 81 749 (16,6 %) aller Actions
- **Erweitertes Suchen:** In OWID gibt es einige Funktionalitäten für erweitertes Suchen. Damit sind wörterbuchübergreifende Suchen und -spezifische Suchen gemeint, die es Benutzer*innen erlauben, Wörterbuchinhalte über mehr als das bloße Stichwort aufzufinden. Wir bezeichnen den Einsatz dieser Möglichkeiten im Folgenden als „Erw. Suche“. Auf diesen Aktionstyp entfallen 60 454 (12,3 %) aller Actions.
- **Wörterbuch-Startseiten aufrufen:** OWID enthält neben der OWID-Hauptseite Startseiten zu den verfügbaren Wörterbüchern. Wir kategorisieren einen Besuch dieser Startseiten im Folgenden unter „WB-Startseite“. Auf diesen Aktionstyp entfallen 48 666 (9,9 %) aller Actions.
- **Stichwortlisten einsehen:** Diese Kategorie umfasst alle Seiten, auf denen weitestgehend unkommentiert Stichwörter aufgelistet sind. Darunter fallen bspw. Stichwortlisten für das Neologismenwörterbuch zu bestimmten Jahrzehnten. Auf diesen Aktionstyp entfallen 42 020 (8,5 %) aller Actions.
- **OWID-Hauptseite aufrufen:** In diese Kategorie fallen alle Besuche der Startseite von OWID. Die Startseiten der einzelnen Wörterbücher fallen nicht darunter. Wir bezeichnen diesen Typ im Folgenden als „Hauptseite“. Auf diesen Aktionstyp entfallen 26 377 (5,4 %) aller Actions.
- **Sonstiges:** Alle Actions, die keinem der anderen Aktionstypen zugeordnet sind, fallen in diese Rubrik. Neben URLs, die nicht zugeordnet werden konn-

Städten aufgehalten hat, sondern es könnte auch den jeweils veränderten Einwahlort des Internetproviders widerspiegeln. Darüber hinaus konnten für einige Zugriffe keine Orte festgestellt werden, was als NA (fehlender Wert) in den Daten vermerkt wird.

ten, zählen hierzu auch alle Subkategorien, die jeweils weniger als 4 % aller Actions auf sich vereinen konnten. Das sind bspw. das Aufrufen von Nebentexten zu den Wörterbüchern und Zugriffe über die (inzwischen nicht mehr vorhandene) Buchstabenleiste. Auf diesen Aktionstyp entfallen 23 904 (4,9 %) aller Actions.

- **Verteilerseiten aufrufen:** Diese Kategorie ähnelt teilweise den Stichwortlisten (s.o.). Sie unterscheidet sich aber von diesen insofern, dass unter „Verteilerseiten“ jene Seiten verstanden werden, die entweder wiederum auf Stichwortlisten verweisen oder Stichwörter in strukturierterer Weise auflisten.¹⁰ Wir bezeichnen diesen Typ im Folgenden als „Verteiler“. Auf diesen Aktionstyp entfallen 20 914 (4,2 %) aller Actions.
- **OWID-„Extras“ aufrufen:** Diese Kategorie umfasst alle Ressourcen, die in OWID^{plus} enthalten sind sowie die Bibliographien OBELEX^{meta} und OBELEX^{dict} und die OWID-Korpussuche. Für viele Angebote in OWID^{plus} können wir mit Matomo nicht genau feststellen, wie die einzelnen Ressourcen genutzt werden, da die Angebote dort eher Webapplikationen als Webseiten sind, und wir uns beim Tracking mit Matomo auf Seitenaufrufe beschränken. Ein Zugriff auf die meisten Ressourcen in OWID^{plus} wird so nur als ein Zugriff protokolliert, egal was in der Ressource selbst dann recherchiert wird. Wir bezeichnen diesen Typ im Folgenden als „Extra“. Auf diesen Aktionstyp entfallen 14 828 (3,0 %) aller Actions.

Aktionstypen sind, wie der Name schon sagt, auf der Ebene der Actions definiert. Für einen Visit mit mehr als einer Action können wir somit Sequenzen und Anteile von Aktionstypen ermitteln. Der in Abschnitt 2.1 angesprochene Visitor aus Kolumbien zeigte bspw. beim ersten Besuch die Sequenz „Hauptseite → Hauptsuche → Artikel → Artikel“.

3.2.2 Zugriffe über die Zeit

Die Zugriffe auf OWID unterliegen ausgeprägten monatlichen Schwankungen (siehe Abbildung 2). Der Monat mit den häufigsten Zugriffen im Untersuchungszeitraum ist der Januar 2019 (32 384 Actions in 5 923 Visits). Jeweils im August sind die wenigsten Zugriffe in den beiden Jahren zu verzeichnen, der Monat mit

¹⁰ Ein Beispiel ist die Verteiler-Seite in *lexiko* zu Jahreszeiten: www.owid.de/wb/lexiko/gruppen/jahreszeiten.html [letzter Zugriff am 18.9.2020]

den wenigsten Zugriffen ist der August 2019 (11 912 Actions in 2 883 Visits). Wir interpretierten dies zunächst so, dass das hauptsächlichliche Publikum von OWID aus dem deutschsprachigen akademischen Bereich stammt und diese Berufsgruppe sich im August im Sommerurlaub befindet. Ob die Mehrzahl der Visits tatsächlich aus einem akademischen Umfeld stammt, können wir anhand der Matomo-Daten nicht verifizieren. Die Herkunftsländer der Visits sind aber enthalten. Aus Deutschland stammen insgesamt 48,4 % aller Visits. Nimmt man Österreich und die Schweiz hinzu, sind es 57,6 %. Schaut man sich den Anteil der Zugriffe aus Deutschland, Österreich und der Schweiz über die Zeit hinweg an, zeigt sich allerdings, dass in den Sommermonaten dieser Anteil eher steigt (den höchsten Wert erreicht er mit 66,9 % im August 2019). Im August 2018 und 2019 werden also eher *weniger* Zugriffe außerhalb Deutschlands, Österreichs und der Schweiz verzeichnet – die „Sommerferieninterpretation“ trägt somit nicht zur Aufklärung dieses Musters bei.

Ungeachtet der starken Schwankungen ist ein sehr schwacher Aufwärtstrend der Zugriffe über die Zeit hinweg zu verzeichnen (Actions: Spearman's $\rho = 0,087$; $p_{\text{perm}} = 0,016$; Visits: $\rho = 0,084$; $p_{\text{perm}} = 0,022$).¹¹ Das Verhältnis zwischen Actions und Visits, also wie viele Actions in einem Visit durchschnittlich stattfinden, schwankt zwischen 4,13 (im August 2019) und 6,78 (im April 2018). Hier ist keine Entwicklung über die Zeit festzustellen. Wir haben also keinen Hinweis darauf, dass Visits über die Zeit hinweg länger oder kürzer werden (gemessen an der Zahl der Actions pro Visit).

11 Der Spearmansche Korrelationskoeffizient wurde anhand eines tagesbasierten Datensatzes berechnet. Mit demselben Datensatz kann auch ein lineares Modell (Regression) berechnet werden. Wie schwach der Effekt ist, zeigt sich am Effektschätzer für die Actions $\beta = 0,195$, der so interpretiert werden kann, dass pro Tag 0,195 Actions hinzukommen. Gegeben dieses Modell, dauert es ca. 5 Tage bis eine einzige Action mehr als zuvor verzeichnet wird. Für die Visits gilt $\beta = 0,023$ (also ca. alle 44 Tage ein zusätzlicher Visit).

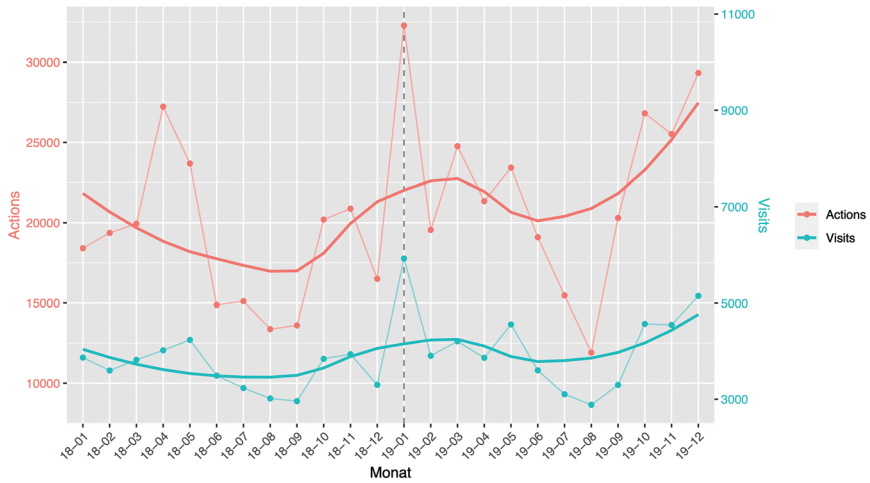


Abbildung 2: Monatliche Zugriffe auf OWID in den Jahren 2018 und 2019, gemessen als Actions und Visits mit jeweiligen LOESS-Anpassungslinien. Bitte beachten Sie die unterschiedlich skalierten y-Achsen für Actions (links) und Visits (rechts). Die gestrichelte Linie gibt den Jahreswechsel an.

Der zeitliche Verlauf der Actions lässt sich noch weiter aufschlüsseln nach den in Abschnitt 3.2.1 eingeführten Aktionstypen. In dieser Aufspaltung entfallen die Visits, da die Aktionstypen nur auf der Ebene der Actions definiert werden können. Abbildung 3 zeigt, dass insbesondere der Aktionstyp „Artikel“ dem Muster für Komplett-OWID sehr ähnlich ist. Das ist nicht überraschend, denn immerhin entfallen über den kompletten Zeitraum hinweg 35,3 % aller Actions auf die Artikel. Die anderen Aktionstypen zeigen diversere Verläufe, und es ist schwer, hier ein Muster auszumachen, das über viele Aktionstypen hinweg konsistent ist. Auffällig sind einzelne „Ausschläge“, wie bspw. der Anstieg des Aktionstyps „Stichwortliste“ im Januar 2019 um mehr als das Dreifache im Vergleich zu den umgebenden Monaten. Dieser ist auf die extensive Berichterstattung über die Liste neuer Einträge aus dem Jahr 2018 im Neologismenwörterbuch zurückzuführen, die im Januar 2019 veröffentlicht wurde. Allein die Zugriffe über Facebook-Links und von der Seite noizz.de (eine Nachrichtenseite für ein jüngeres Publikum des Springer-Verlags) sorgen in diesem Monat für über die Hälfte aller Actions für alle OWID-Stichwortlisten und treten in den Monaten zuvor fast überhaupt nicht bei den auf OWID verweisenden Seiten in Erscheinung.¹² Ähnliches gilt in weitaus

¹² Diese Informationen können wir dem Referrer-Feld entnehmen, in der – falls vorhanden – die Webseite enthalten ist, die auf OWID verwiesen hat.

geringerem Maße für den Anstieg des Aktionstyps „Extra“ von September bis November 2019, der höchstwahrscheinlich erklärbar ist durch das offizielle Release der Ressource „Lexik des gesprochenen Deutsch“ (LeGeDe), das mit einem Workshop am IDS Ende September 2019 verbunden war. Für diese temporären Anstiege sind also singuläre Ereignisse verantwortlich, was eine Systematisierung im Gesamtzusammenhang erschwert. Ähnliches beobachteten wir auch bei unseren Analysen von Log-Files des deutschen Wiktionary (Müller-Spitzer et al. 2015), wo wir starke Einflüsse von temporär sozial relevanten Ereignissen (bspw. politischen Diskussionen, Quizshows und Fußballspielen) auf die Nachschlagehäufigkeiten zeigen konnten.

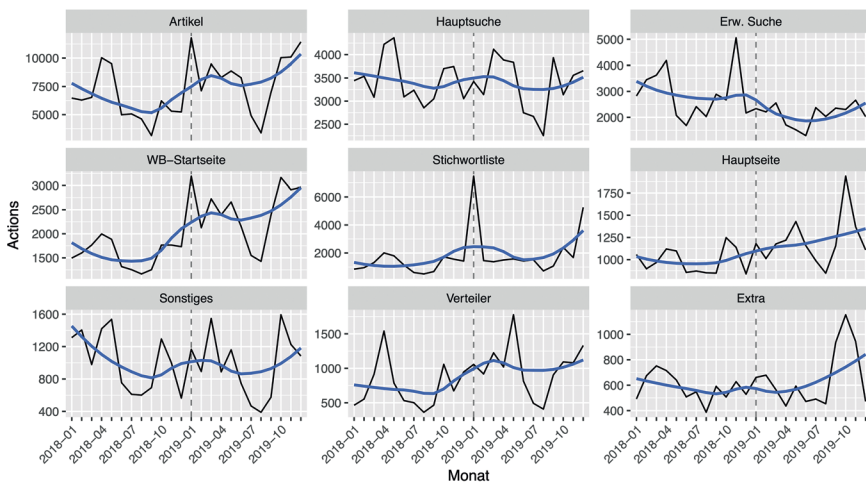


Abbildung 3: Monatliche Häufigkeit verschiedener Aktionstypen auf OWID in den Jahren 2018 und 2019 mit jeweiligen LOESS-Anpassungslinien. Bitte beachten Sie die unterschiedlich skalierten y-Achsen. Die gestrichelten Linien geben den Jahreswechsel an.

3.2.3 Nutzungsmuster als Aktionstyp-Sequenzen

Bisher haben wir einzelne Actions in Aktionstypen kategorisiert und im Zeitverlauf analysiert. Hierbei ist stets die einzelne Action die Analyseeinheit. Das Nachschlagen in einem Wörterbuch ist damit allerdings nicht hinreichend beschrieben, denn eine Konsultationshandlung kann deutlich länger sein als das Aufrufen eines Artikels. Die Matomo-Daten zu OWID unterstützen diese Annahme. Zwar umfassen ziemlich genau die Hälfte (49,9 %, $n = 46\ 353$) aller Visits tatsäch-

lich nur eine Action, der Mittelwert über den gesamten Betrachtungszeitraum hinweg liegt jedoch bei 5,5 Actions pro Visit. Es scheint sich also ein Blick zu lohnen, welche Actions oft nacheinander ausgeführt werden.

Hierzu abstrahieren wir von der Ebene der Visits, denn das, was wir mit dem Begriff „Nutzungsmuster“ erfassen möchten, sind Sequenzen von Actions, die nicht unbedingt identisch sein müssen mit dem kompletten Visit. Eine erste einfache Annäherung sind Abfolgen von Aktionstypen, die sich anhand der Matomdaten erstellen lassen. So können wir besonders häufige Sequenzen identifizieren. Aus dem in Abschnitt 2.1 bereits zitierten Zugriff aus Kolumbien, der die Aktionstypsequenz „Hauptseite → Hauptsuche → Artikel → Artikel“ aufwies, ergeben sich bspw. die folgenden fünf Zweier-Sequenzen von Aktionstypen:

- #Start# → Hauptseite
- Hauptseite → Hauptsuche,
- Hauptsuche → Artikel
- Artikel → Artikel
- Artikel → #Ende#.

Die Codes für den Beginn (#Start#) und das Ende (#Ende#) des Visits fügen wir ein, um die Information, womit ein Visit begonnen und beendet wurde, nicht zu verlieren. Die Länge der Sequenzen kann variiert werden. Würden wir bspw. Vierer-Sequenzen betrachten, würden für das obige Beispiel die folgenden drei Vierer-Sequenzen anfallen:

- #Start# → Hauptseite → Hauptsuche → Artikel
- Hauptseite → Hauptsuche → Artikel → Artikel
- Hauptsuche → Artikel → Artikel → #Ende#

Wir schließen zunächst Sequenzen von bis zu fünf Aktionen in unsere Betrachtungen ein. Die Anzahl der in die Analyse einbezogenen Visits sinkt mit der Fenstergröße der Sequenzen, denn der Visit muss mindestens so lang sein wie die Fenstergröße. Wie das die Anzahl und den Anteil einbezogener Visits beeinflusst, kann in Tabelle 1 abgelesen werden. Würden wir bspw. auch Sequenzen von zehn Aktionen betrachten, würden wir nur noch 11,9% aller Visits überhaupt in die Betrachtungen mit einfließen lassen.

Tabelle 1: Übersicht über die Anzahl der Visits, den prozentualen Anteil an allen Visits sowie die Anzahl der n-Gramme, die in die Berechnung der n-Gramme unterschiedlicher Fenstergröße eingingen. N-Gramme der Größe 10 sind nur zu Vergleichszwecken aufgeführt und werden im Folgenden nicht mehr beachtet.

| Sequenzlänge | Anzahl Visits | % aller Visits | Anzahl Sequenzen |
|--------------|---------------|----------------|------------------|
| 2 | 46 171 | 49,7 % | 492 464 |
| 3 | 34 235 | 36,9 % | 422 421 |
| 4 | 26 473 | 28,5 % | 372 662 |
| 5 | 21 906 | 23,6 % | 337 055 |
| 10 | 11 017 | 11,9 % | 232 752 |

Anhand der Aktionstyp-Sequenzen lassen sich verschiedene Fragen untersuchen. Zweiersequenzen eignen sich beispielsweise dazu, die häufigsten Eintrittsseiten in OWID zu extrahieren. Hierzu können wir alle Zweiersequenzen selektieren, deren erstes Element #Start# ist. Am häufigsten ($n = 11\,000$) ist hier die Sequenz #Start# → WB-Startseite. Die häufigste Einstiegsseite ist also nicht die OWID-Hauptseite, sondern der Einstieg direkt auf der Startseite eines bestimmten Wörterbuchs. Allerdings liegt die entsprechende Zweiersequenz #Start# → Hauptseite mit $n = 10\,872$ knapp dahinter auf dem zweiten Rang der Einstiegsaktionen. Trotzdem ist es für uns ein interessanter Hinweis, dass sowohl einzelne Wörterbücher gezielt aufgerufen werden als auch das gesamte Portal. Innerhalb aller Zweiersequenzen rangieren diese beiden Einstiegssequenzen allerdings nur auf den Plätzen 10 und 11. Abbildung 4 zeigt, dass das Aufrufen zweier Artikel hintereinander die mit Abstand häufigste Zweiersequenz ist. Wiederholte erweiterte Suchen befinden sich auf Rang 2, wiederholte Hauptsuchen finden sich auf Rang 6. Einige klassische Zugriffsstrukturen finden sich ebenfalls in den häufigsten Zweiersequenzen: so bspw. der Zugriff über Stichwortlisten (Rang 3) oder die Hauptsuche (Rang 4). Interessanterweise befindet sich der Artikelzugriff über *erweiterte* Suchen erst auf Rang 38 der häufigsten Zweiersequenzen ($n = 2667$; 0,54 %).

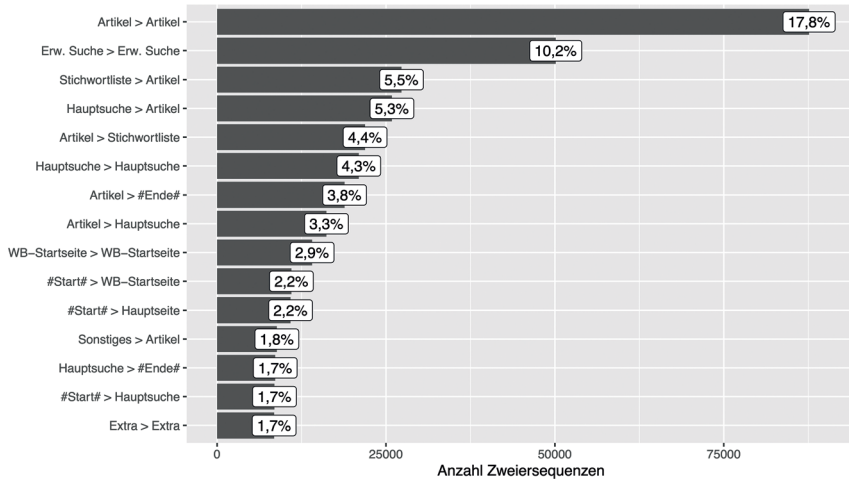


Abbildung 4: Die 15 häufigsten Zweiersequenzen in Visits mit mindestens zwei Actions. Die Prozentangaben beziehen sich auf den Anteil der jeweiligen Sequenz an allen Zweiersequenzen in Visits mit mindestens zwei Actions.

Will man Artikel-Zugänge über die Hauptsuche und die erweiterten Suchen vergleichen, muss man allerdings auch die Grundwahrscheinlichkeiten beachten, wie oft *insgesamt* die Hauptsuche und die erweiterten Suchen verwendet werden. Sie sind allerdings nahezu gleichauf (60 264 vs. 59 324), was dafürspricht, dass es tatsächlich viel wahrscheinlicher ist, dass Visitors einen Zugang zu Artikeln über die Hauptsuche finden als über die erweiterten Suchen. Es ist anhand der von Matomo gesammelten Daten nicht herauszufinden, woran das liegt. Es könnte tatsächlich sein, dass die erweiterten Suchen den Visitors nicht gut genug dabei helfen, den geeigneten Artikel zu finden. Es kann aber auch sein, dass die erweiterten Suchen dazu eingesetzt werden, Wörter über bestimmte Eingrenzungskriterien zu finden (bspw. Neologismen aus Mitte der 90er-Jahre, die Präverbfügungen darstellen und als stilistisch „derb“ markiert sind), zu denen dann aber nicht der entsprechende Artikel gelesen wird. Das Ziel dieser Benutzungshandlung wäre dann das Auffinden eines Worts nach bestimmten Kriterien und nicht die weitergehenden Informationen zu dem Wort, die erst im Wörterbuchartikel selbst zu finden wären.

Betrachten wir Dreier-, Vierer- und Fünfersequenzen (Abbildung 5 zeigt die 15 häufigsten Fünfersequenzen, die besprochenen Muster sind aber für die Fenstergrößen 3 und 4 äquivalent), fällt auf, dass sich einige der Muster, die für die Zweiersequenzen (Abbildung 4) zu sehen sind, für größere Fenster fortsetzen: Platz 1 ist von ununterbrochenen Abfolgen von Artikelaufrufen belegt, auf Platz 2 gilt

dasselbe durchgehend für erweiterte Suchen. Sich stets abwechselnde Folgen von Artikel- und Stichwortlistenaufrufen finden sich auf den Plätzen 3 und 4. Die beiden folgenden Ränge werden von abwechselnden Sequenzen zwischen Artikeln und Hauptsuchen eingenommen.

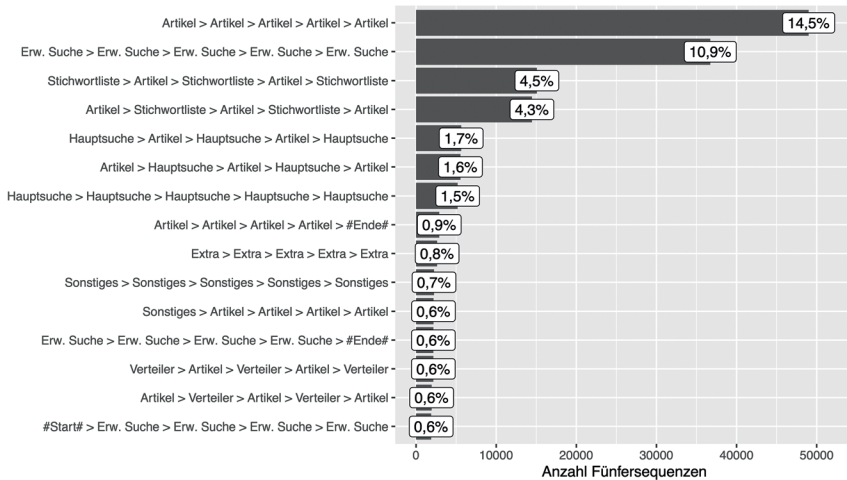


Abbildung 5: Die 15 häufigsten Fünfersequenzen in Visits mit mindestens fünf Actions. Die Prozentangaben beziehen sich auf den Anteil der jeweiligen Sequenz an allen Fünfersequenzen in Visits mit mindestens fünf Actions.

Gesondert möchten wir auf die Ränge 9 und 10 hinweisen, die jeweils durchgehende Sequenzen des Aktionstyps „Extra“ und „Sonstiges“ sind. Im Gegensatz zu Verteilern, Stichwortlisten und den verschiedenen Suchtypen tauchen „Extras“ und „Sonstiges“ nicht gemeinsam mit Artikelaufrufen in den häufigsten Sequenzen auf, wohl aber als ununterbrochene Sequenzen mit sich selbst. Das legt den Verdacht nahe, dass die Visitors von OWID tendenziell *entweder* nach Wörtern oder Wortartikeln recherchieren und dabei Wortlisten oder Suchen verwenden *oder* sich für bestimmte Teilressourcen interessieren, die typischerweise nicht zum Kernbestand eines Wörterbuchs gehören (zu ‘outer texts’ und ‘outer features’ von Wörterbüchern siehe Klosa & Gouws 2015).

3.2.4 Visitor-Typen

Dies legt den Verdacht nahe, dass es unterschiedliche Typen von Visitors gibt, die auf OWID zugreifen. Die bloße Betrachtung der unterschiedlichen Aktionssequenzen bringt uns hier inhaltlich allerdings nur begrenzt weiter (allenfalls als Anregungen zu „Verdachtsmomenten“ wie oben beschrieben). Wenn wir uns für verschiedene Visitor-Typen interessieren, können wir eher die Methode der „Personas“ (vgl. Cooper, Reimann & Cronin 2007: 77ff) aufgreifen, die ursprünglich aus dem Bereich der Softwareentwicklung stammt und unserem Wissensstand nach noch nicht für die Nutzungsforschung zu Online-Wörterbüchern eingesetzt wurde.

Personas sind Modelle von Nutzer*innen (hier eines Wörterbuchportals, ursprünglich, wie erwähnt, von Softwareprodukten). Cooper et al. (2007: 76) beschreiben sie als „composite archetypes“, die auf beobachteten Verhaltensdaten beruhen, die in einer Modellierungsphase so formalisiert werden, dass die jeweilige Persona über Interaktionsmuster mit einem System identifiziert werden kann. Personas können also als operationalisierbare Prototypen potentieller Nutzer*innen eines Systems aufgefasst werden. Eigentlich stehen dabei hinter den einzelnen Personas Beschreibungen von Motivationen und Zielen, die bspw. anhand von Interviews mit Nutzer*innen, Expert*innen, Fokusgruppen oder aus Modellen der Marktsegmentierung gewonnen werden. Typischerweise wird jeder Persona sogar ein Bild zugeordnet.

Wir möchten ausdrücklich darauf hinweisen, dass wir in diesem Beitrag den Personas-Ansatz nicht in dieser Breite ausarbeiten können. Wir wollen an dieser Stelle zunächst ausprobieren, wie weit ein solcher Ansatz im Kontext der Benutzungsforschung für ein Wörterbuchportal überhaupt trägt, und wie viele Visitors wir mit einigen einfachen Personas erfolgreich kategorisieren können. Im Gegensatz zu den oben beschriebenen Wegen zur Erstellung von Personas werden wir hier Operationalisierungen (also Kriterien, die wir ansetzen, damit ein Visitor in eine bestimmte Kategorie fällt) verwenden, die uns erstens sinnvoll erscheinen und sich zweitens gegenseitig ausschließen.

Grundlage der folgenden Analyse ist ein Datensatz, in dem ein Visitor als ein Fall behandelt wird. Für jeden Visitor berechnen wir – über alle Visits hinweg – wie viel Prozent der Actions auf die verschiedenen Aktionstypen entfallen. Wir nehmen zunächst vier potentielle Personas an, die wir folgendermaßen bezeichnen und operationalisieren:

1. Die „klassischen Visitors“: Die klassischen Visitors lesen hauptsächlich die Artikel in den verschiedenen Wörterbüchern, die auf OWID zur Verfügung stehen. Dabei nutzen sie die „klassischen“ Zugriffsstrukturen wie Stichwortlisten, Verteilerseiten und die Hauptseiten (sowohl die OWID-Hauptsei-

te als auch die Hauptseiten der verschiedenen Wörterbücher). Diese Gruppe „stöbert“ auch gerne einmal im Wörterbuch bzw. vertieft sich in Inhalte und lässt sich von Artikel zu Artikel leiten. Diese Gruppe operationalisieren wir dadurch, dass mindestens 75 % aller Actions pro Visitor auf Artikel, Stichwortlisten oder Verteilerseiten und die verschiedenen Hauptseiten entfallen.

2. Die „Sucher*innen“: Die Sucher*innen suchen nach einer ganz bestimmten Information und setzen dabei überwiegend die OWID-Hauptsuche ein. Diese Gruppe operationalisieren wir dadurch, dass mindestens 50 % aller Actions pro Visitor auf die OWID-Hauptsuche entfallen.
3. Die „Power-Visitors“: Die Power-Visitors sind ebenfalls auf der Suche nach ganz bestimmten Informationen, die wortübergreifend sind und auf systematische Aspekte der Sprache abzielen. Sie setzen dabei die Möglichkeiten ein, die ihnen die erweiterten Suchfunktionalitäten bieten. Diese Gruppe identifizieren wir dadurch, dass mindestens 50 % aller Actions pro Visitor auf die erweiterten Suchen entfallen.
4. Die „Special-Interest-Gruppe“: Die Special-Interest-Gruppe interessiert sich weniger für die Wörterbuchartikel, die auf OWID enthalten sind, sondern eher für die Dinge, die abseits der enthaltenen Wörterbücher angeboten werden. Sie nutzen Ressourcen wie die Korpussuche, OWID^{plus} oder die Umtexte der Wörterbücher. Diese Gruppe identifizieren wir dadurch, dass mindestens 50 % aller Actions pro Visitor auf die Aktionstypen „Extra“ und „Sonstiges“ entfallen.
5. Andere: Dieser Gruppe werden alle Visitors zugeordnet, die in keine der anderen Gruppen fallen.

Auch wenn es theoretisch möglich ist, dass ein Visitor bspw. sowohl in Gruppe 3 und Gruppe 4 fallen könnte (wenn nämlich exakt gleich viele Actions auf erweiterte Suchen und Extras oder Sonstiges entfallen und sonst keine Actions vorhanden sind), gibt es keinen solchen Überschneidungsfall in unseren Daten.

Damit die oben genannten Gruppenberechnungen überhaupt Sinn ergeben, können wir nur Visitors beachten, für die mindestens vier Actions verbucht sind.¹³ Dies hat zur Folge, dass wir von den insgesamt 45 784 Visitors noch 17 077 Visitors (37 %) beachten können. Diese 17 077 Visitors sind aber für 92 % aller Actions

¹³ Gruppe 1 ist über einen Mindestprozentsatz von 75 % der für diese Gruppe relevanten Actions definiert. Erst ab vier Actions ist dies eine sinnvolle Mindestgrenze, da in diesem Fall drei der vier Actions auf die Kriterien zutreffen müssen.

verantwortlich. Wir verlieren durch die Einschränkung auf Visitors mit mindestens vier Actions also relativ viele Visitors für die Auswertung, aber relativ wenige Actions.

Abbildung 6 zeigt die Verteilung der analysierten Visitors auf die fünf Gruppen. Über die oben definierten Operationalisierungen können wir 23,5 % der Visitors nicht zuordnen. Wir können also die deutliche Mehrheit (76,5 %) der analysierten Visitors einer der vier anderen Gruppen zuordnen.

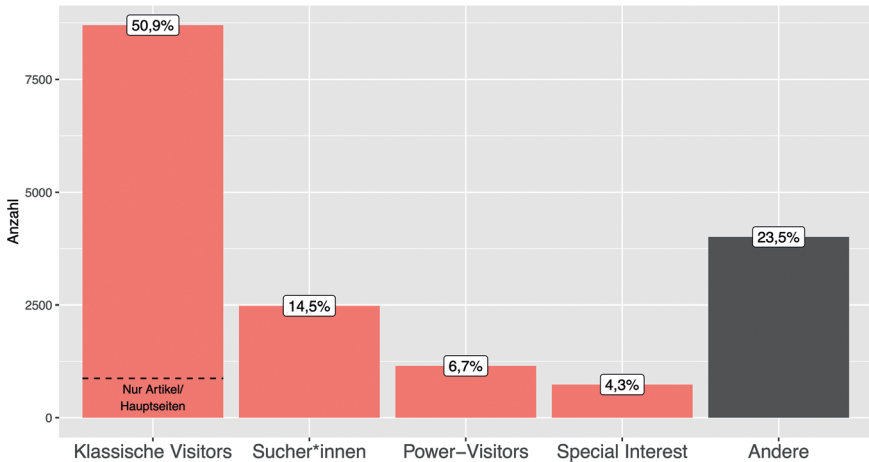


Abbildung 6: Verteilung der analysierten Visitors auf die vier definierten Personas und „Andere“. Die Prozentangaben beziehen sich auf die Gesamtzahl der in die Analyse einbezogenen Visitors mit mindestens vier Actions. Die gestrichelte Linie bei den Klassischen Visitors gibt die Anzahl derer an, die nur Artikel und Hauptseiten besuchten ($n = 873$; 5,1% aller hier analysierten Visitors).

Die Mehrheit der analysierten Visitors (50,9 %) wird der Gruppe der „klassischen Visitors“ zugeordnet, die momentan auch noch recht breit definiert ist. Hier wäre eine Verfeinerung dahingehend möglich, Visitors, die eher auf Stichwortlisten und Verteilerseiten zugreifen, von jenen abzugrenzen, die nur Artikel und Hauptseiten aufrufen. Die letztgenannte Subgruppe von Gruppe 1 benutzt also – wenn überhaupt – ausschließlich die Hauptseiten von OWID oder den einzelnen Wörterbüchern als Zugriffsstrukturen auf Artikel. In Abbildung 6 haben wir diese Subgruppe durch eine gestrichelte Linie abgetrennt. Es sind 873 Visitors, was innerhalb der Gruppe der „klassischen Visitors“ 10 % ausmacht und über alle hier analysierten Visitors 5,1%. 2 482 (14,5 %) Visitors fallen in die Gruppe der Su-

cher*innen. Erwartungsgemäß gibt es mehr von diesen häufig die Hauptsuche einsetzenden Visitors als jene, die wir als „Power-Visitors“ bezeichnen ($n = 1151$, 6,7%). 736 Visitors fallen in die „Special-Interest-Gruppe“ – betrachten also vorwiegend bspw. Umtexte, nutzen die OWID-Korpussuche oder verwenden Ressourcen auf OWID^{plus}. Zum Vergleich: Das sind weniger Visitors als die Subgruppe der klassischen Visitors, die ausschließlich Hauptseiten und Artikel aufrufen (gestrichelte Linie in Abbildung 6).

Eine solche Herangehensweise kann uns zeigen, wie sich die Visitors unseres Wörterbuchportals auf verschiedene vordefinierte Gruppen verteilen. Dies könnte u. a. dazu eingesetzt werden, das Portal als Ganzes stärker auf bestimmte Gruppen hin auszurichten oder Mittel und Wege zu finden, um bestimmte Zugriffsstrukturen und Inhalte prominenter darzustellen.

In einer früheren Version dieses Beitrags hatten wir versucht, die Verweildauern auf einzelnen Seiten einer Analyse zu unterziehen. Solche Analysen haben jedoch grundlegende Probleme, von denen wir nur einige nennen möchten. Erstens fallen extrem viele Verweildauern weg. Wir haben in Abschnitt 2.1 auf den Grund hingewiesen, nämlich dass für die letzte Action in einem Visit aus technischen Gründen keine Verweildauer angegeben ist. Besteht ein Visit aus nur einer Action, haben wir also gar keine Information zur Verweildauer. Das zweite Problem ist ein konzeptuelles: Verweildauern können fehlinterpretiert werden. So können Verweildauern bspw. künstlich hoch sein, wenn eine Webseite im Hintergrund oder in einem anderen Browser-Tab geöffnet ist. Verweildauer hat also nur mittelbar etwas mit Lesedauer zu tun – und wie eng der Zusammenhang ist, können wir anhand der anfallenden Daten nicht einschätzen. Drittens können unterschiedliche Benutzungsszenarien die Verweildauer beeinflussen: Beim ersten Lesen eines Wörterbuchartikels ist die Verweildauer vermutlich deutlich höher als bei einem Zugriff, bei dem bereits bekannte Information lediglich wiedergefunden werden muss, um sie z.B. zu kopieren. Wir haben die auf den ersten Blick attraktive Möglichkeit der Analyse von Verweildauern aus diesen Gründen nicht in die finale Version des Beitrags aufgenommen.

3.3 Diskussion

In den in Abschnitt 3.2 dargestellten Auswertungen haben wir bewusst auf längere Einordnungen und Bewertungen verzichtet. Die Potentiale und Limitationen der Analysen werden wir konzentriert in den folgenden Abschnitten darstellen. Damit behandeln wir auch die ersten beiden Fragestellungen:

- Welche neuen Erkenntnisse können wir aus Nutzungsdaten zu linguistischen Online-Ressourcen gewinnen, die uns dabei helfen können, diese Ressourcen

benutzerfreundlicher, d.h. zu besseren Werkzeugen zum Lösen linguistischer Problemstellungen, zu machen?

- Welcher Limitationen der anfallenden Datenspuren müssen wir uns bei der Interpretation von so erhobenen Daten bewusst sein?

3.3.1 Potentiale

Allein der Umfang der während der Benutzung anfallenden Daten bietet ein bestimmtes Potential – immerhin haben wir es in den zwei von uns untersuchten Jahren mit fast einer halben Million Zugriffe auf OWID zu tun. Diese Zugriffe in Aktionstypen zu kategorisieren scheint uns ein gangbares Vorgehen, um die Datenmenge in den Griff zu bekommen und um von Einzel-URLs zu abstrahieren, denn im analysierten Datensatz sind ca. 125 000 verschiedene URLs enthalten. Wir haben hierfür Aktionstypen gewählt, die uns für unser Wörterbuchportal nützlich erschienen, aber eine solche Kategorisierung der aufgerufenen URLs ist natürlich auf andere linguistische Online-Ressourcen anpassbar. Wir hätten auch andere, beispielsweise feingliedrigere, Kategorisierungen wählen können: Unter einer entsprechenden Fragestellung könnten wir die Artikelaufrufe nach bestimmten Wörterbüchern aufteilen, um z.B. vergleichen zu können, wie oft auf die Artikel des Neologismenwörterbuchs im Vergleich zu den verschiedenen Diskurswörterbüchern in OWID zugegriffen wurde. Da wir eine funktional orientierte Kategorisierung der Aktionstypen vorgenommen haben, können wir darüber hinaus weitere Vergleiche anstellen. So können wir u.a. sehen, dass die Hauptsuche häufiger verwendet wird als die erweiterten Suchen (rund 82 000 vs. rund 60 000 Actions), aber auch, dass insgesamt deutlich häufiger gesucht als über Wortlisten zugegriffen wird (rund 142 000 vs. rund 63 000 Actions). Als Vorgriff auf die im nächsten Abschnitt besprochenen Limitationen dieses Ansatzes muss aber bereits hier konstatiert werden, dass solche Vergleiche den globalen Aufbau einer Internet-Ressource außer Acht lassen. Wenn bspw. Suchmöglichkeiten einfach viel prominenter auf der Seite präsentiert werden als Zugriffsmöglichkeiten über Wortlisten, kann diese Verteilung auch mehr die Seitenarchitektur als Präferenzen der Benutzer*innen spiegeln.

In der Analyse der Zugriffshäufigkeit über die Zeit war auffällig, dass generelle Trends (z.B. ein Zuwachs von Zugriffen über die Zeit hinweg) hinter dem Einfluss singulärer Ereignisse zurückzutreten scheinen. Zwar kann ein genereller Trend quantitativ-statistisch gezeigt werden, viel deutlicher sind aber einzelne „Ausschläge“ oder „Peaks“ im Zeitverlauf. Diese können – und darin sehen wir ein Potential dieser Kennzahlen – herangezogen werden, um die Wirksamkeit von Disseminierungsmaßnahmen wissenschaftlicher Inhalte in die Öffentlichkeit zu

evaluieren. Wir führten hier ein Beispiel an, wo alleine die Zugriffe über Facebook-Links und von der Seite noizz.de über die Hälfte aller Zugriffe auf Stichwortlisten im Januar 2019 ausmachten. Hier ließe sich bspw. über eine gezielte Auswertung des Referrer-Felds, also die verweisende Webseite, die auf OWID verlinkt, gut nachzeichnen, auf welchen Plattformen es sich besonders lohnt, Inhalte zu posten. Variiert man gezielt die Verbreitungsmaßnahme (bspw. Bewerbung der Inhalte am Wochenende vs. Mitte der Woche) könnte man auch evaluieren, welche Maßnahme vielversprechender ist. Eine solche Auswertung planen wir für zukünftige Analysen.

Die Analyse von Nutzungssequenzen erlaubt uns auch einen Einblick in generalisierte Verhaltensmuster der Visitors auf OWID. Hier lassen sich bspw. gezielt Einstiegssequenzen identifizieren, also jene Seiten, die als erstes aufgerufen werden, wenn ein Visitor OWID aufruft. Wir können hier beobachten, dass der Einstieg über die Startseiten von einzelnen Wörterbüchern ebenso wahrscheinlich ist wie der Einstieg über die OWID-Hauptseite. Für uns bedeutet das, dass die Konzeption von OWID als Wörterbuchportal, in dem mehrere Wörterbücher unter Kenntlichmachung ihrer konzeptionellen Eigenständigkeit unter einem Dach vereint werden, aufrechterhalten werden sollte. Wenn die Wörterbuch-Startseiten als Einstiegsseiten ebenso relevant sind wie die OWID-Hauptseite, spricht das gegen eine Konzeption von OWID, in der alle Wortartikel in einer großen Masse aufgehen und die Zuordnung zu einem ganz bestimmten Wörterbuch nur mehr am Rande relevant ist (zu Wörterbuchportalen allgemein s. auch Engelberg & Müller-Spitzer 2013). Ferner ist das Ergebnis, dass ununterbrochene Sequenzen von Artikelaufrufen mit Abstand die häufigste Nutzungssequenz sind, für uns eher überraschend. Ausgehend von gedruckten Wörterbüchern hätte man auch erwarten können, dass zwischen den Artikelaufrufen Stichwortlisten oder andere Verteilerseiten besucht werden. Hier wäre es interessant zu erfahren, wie die Nutzungsmuster bei anderen Wörterbuchportalen ausfallen.

Der Ansatz, verschiedene Visitor-Typen über Personas zu identifizieren, scheint uns insofern vielversprechend, als sich über unsere Ad-hoc-Operationalisierungen die Mehrheit der OWID-Visitors kategorisieren lassen (zumindest jene, für die genug Actions vorliegen). Auch die Verteilung auf die verschiedenen Gruppen kann man sinnvoll interpretieren: So gibt es bspw. mehr Sucher*innen als Power-Visitors und die Special-Interest-Gruppe ist die kleinste. Um den Personas-Ansatz noch fruchtbarer für die Wörterbuchbenutzungsforschung zu machen, müssten wir insbesondere weiter an den Operationalisierungen der einzelnen Personas arbeiten. Diese sind bei uns zwar inhaltlich motiviert, doch die konkreten numerischen Kriterien sind arbiträr festgelegt. Hier müsste, um diesen Ansatz fortzuentwickeln, eine externe Motivation dieser Grenzwerte hinzutreten. Dies kann beispielsweise über vorangestellte Befragungen (vgl. Cooper

et al. 2007: 80f) geschehen, die die Personas stärker motivieren und mit mehr „Leben“ füllen.

3.3.2 Limitationen

Einige Einschränkungen entstehen bereits bevor die Daten überhaupt analysiert werden können. Zunächst müssen wir diverse Datenverluste in Kauf nehmen. An mehreren Stellen in diesem Beitrag ist dies relevant: angefangen bei der Dunkelziffer nicht trackbarer Visitors, über die Selektion von Visits mit einer Mindestzahl von Actions für die Sequenzanalysen bis hin zum Ausschluss von Visitors, die nicht mindestens vier Actions verzeichneten, da ansonsten keine sinnvolle Einordnung in eine Gruppe möglich ist. Jeder dieser Datenverluste ist u.E. sinnvoll zu begründen. Eine Limitation der Aussagekraft entsteht aber dadurch, dass die Verluste an einigen Stellen additiv wirken. So ist die Dunkelziffer nicht getrackter Visitors allen weiteren Selektionen und anschließenden Analysen vorgeschaltet und unumgänglich.

Eine weitere Limitation ist ebenfalls den Analysen vorgeschaltet, nämlich die Abhängigkeit der Action-Zählung von der Seitenarchitektur, also welche Inhalte auf wie viele HTML-Seiten verteilt werden. Präsentieren wir beispielsweise in einer Variante alle Lesarten eines Worts auf einer URL und in einer anderen Variante auf mehreren URLs¹⁴, muss die zweite Variante mehr Actions im Datensatz generieren, obwohl die gleichen Inhalte vermittelt werden. Doch selbst wenn im Browser der Besucher*innen dasselbe „geschieht“, kann die dahinterliegende Architektur so variieren, dass es unter Umständen starken Einfluss auf die Zählung von Actions hat. Das gilt auch für zwei erweiterte Suchen in OWID. Konzeptionell sind sie gleich aufgebaut, technisch aber unterschiedlich realisiert. So wird die Seite für die erweiterte Suche von *ellexiko* (<https://www.owid.de/suche/ellex/erweitert>) mit jeder Suche neu auf dem Server generiert (und von Matomo als Action

14 In der Anfangszeit von OWID waren die Artikel von *ellexiko* und dem Neologismenwörterbuch so organisiert, dass es eine Einstiegsseite für jeden Artikel gab und von dieser verlinkt für jede Bedeutung/Lesart je eine weitere URL (z. B. „Unglück“ in der Lesart „Leid“: <https://www.owid.de/artikel/253983/Leid>). Für Artikel mit vielen Lesarten mag diese Organisationsform sinnvoll erscheinen, da die Textmenge übersichtlicher präsentiert werden kann, wenn man sie auf mehrere Seiten verteilt. Um aber an die Informationen zu einer Lesart zu gelangen, sind Benutzer*innen gezwungen, wenigstens zweimal einen Link aufzurufen. Da die überwiegende Anzahl der Stichwörter im Neologismenwörterbuch nur eine Lesart besitzt, wurde beschlossen, die Informationen zur Lesart gleich auf der Einstiegsseite zu präsentieren und auf diese Weise den Benutzer*innen den zweiten Klick zu ersparen. Das bedeutet, dass in dem einen Wörterbuch für das Auffinden der gleichen Information zwei Actions, bei dem anderen nur eine Action verzeichnet wird.

protokolliert), selbst dann, wenn der Benutzer in der Ergebnisliste die zweite, dritte oder letzte Seite auswählt. Die Seite für die erweiterte Neologismensuche (<https://www.owid.de/docs/neo/suche>; seit Oktober 2019) hingegen wird nur einmal vom Server ausgeliefert, womit nur eine Action anfällt. Jede Zustandsveränderung im Browser danach, bspw. neue Suchen, Blättern im Suchergebnis, Filtern und Highlighting erfolgt dynamisch, ohne dass die Seite erneut vom Server angefordert wird.¹⁵ Ein Vergleich der Actions für die beiden URLs ist aufgrund dieser unterschiedlichen Implementierung nicht sinnvoll möglich.

Wie erwähnt, sind die vorgenannten Limitationen den eigentlichen Analysen von anfallenden Benutzungsdaten vorgeschaltet, doch auch die Analysen selbst sind bestimmten Beschränkungen unterworfen. Dies gilt auch für die Evaluation von Wissenschaftskommunikation, die wir oben noch als interessantes Potential dargestellt haben. Zwar kann das Referrer-Feld, also die Information, welche URL auf OWID verwiesen hat, in vielen Fällen als zusätzliches Analyseinstrument hinzugezogen werden, doch muss hier beachtet werden, dass dieses Feld lediglich in 44,6 % aller Visits ($n = 41\,468$) überhaupt belegt ist. Liegt uns die Information über die verweisende URL nicht vor, gleichen Wirksamkeitseinschätzungen von Disseminierungsmaßnahmen eher einem „Indizienprozess“: Wir wissen, wann wir Informationen verbreitet haben, und wir können in einem gewissen zeitlichen Abstand evtl. einen Anstieg der Zugriffe auf bestimmte Ressourcen feststellen. Ob diese beiden Umstände allerdings kausal in Verbindung stehen, bleibt unklar. Versucht man einmal, die umgekehrte Perspektive einzunehmen, wird das vielleicht noch klarer: Es fällt nämlich durchaus schwer, für jeden einzelnen sprunghaften Anstieg von Zugriffszahlen auf bestimmte Ressourcen einen konkreten Grund auszumachen (übrigens gilt das auch für die Untersuchungen zum Zusammenhang temporärer sozialer Relevanz und Wiktionary-Zugriffen, die Müller-Spitzer et al. 2015 präsentieren). Bei kausalen Erklärungen von temporären Zugriffsanstiegen sind wir stets auf die Plausibilität bestimmter Verdachtsmomente angewiesen.

15 Prinzipiell lassen sich auch für solche *Single-Page-Apps* (SPA) Tracking-Lösungen mit Matomo realisieren, denn selbst wenn sich die URL nicht ändert, können vom Browser generierte Ereignisse wie z. B. Klicks genutzt werden, um sie mit Matomo-Actions zu assoziieren. Dazu muss aber Anwendungs- und Trackingcode weitgehend vernetzt werden. Die Nachteile solcher Softwareabhängigkeiten sind sorgsam gegen den tatsächlichen Nutzen der Nutzungsverfolgung abzuwägen. Da SPAs innerhalb von OWID eine noch relativ junge und wenig eingesetzte Technologie darstellen, beschränkt sich das Tracking bisher allein auf URLs, die Nutzung der SPAs wird also momentan nicht getrackt. Unabhängig von der technischen Realisierung des SPA-Trackings bleibt das generelle Problem bestehen, dass beispielsweise Nutzungshandlungen auf Suchseite A mit jenen auf Suchseite B schwer verglichen werden können, wenn das Design von Benutzerführung und Interaktionsmöglichkeiten grundlegend verschieden ist.

Für die Weiterentwicklung des Wörterbuchportals selbst schätzen wir den Nutzen einer Zugriffsanalyse über die Zeit eher als beschränkt ein. Ausnahmen könnten hier gezielt ausgewählte Ereignisse sein, die einen Umbau der Plattform betreffen. Stellen wir uns beispielsweise vor, dass eine erweiterte Suche in ihrer Funktionalität zu einem bestimmten Zeitpunkt verändert wird. Nun könnte man über einen gezielten Vergleich zwischen den Zeiträumen vor vs. nach der Änderung überprüfen, ob diese Veränderung von den Nutzer*innen angenommen wird. Die Erwartung wäre, dass sich dies in einer vermehrten oder zielführenderen Nutzung niederschlägt. Letztere würde anhand einer vergleichenden (Vorher- vs. Nachher-Version) Sequenzanalyse sichtbar. Doch auch in einem solchen Szenario sind andere Untersuchungen, nämlich A-B-Testungen, auf die wir noch eingehen werden, zu bevorzugen, da die (evtl. minimalen) Veränderungen, die auf einen Umbau zurückzuführen sind, leicht in den sehr starken Schwankungen, die in Abbildung 2 und Abbildung 3 zu sehen sind, untergehen können.

Einige der Limitationen, derer wir uns bei der „Personas“-Auswertung bewusst sein müssen, haben wir bereits im entsprechenden Abschnitt angesprochen – insbesondere die eher als probeweise zu verstehenden Operationalisierungen der Gruppen. Für diese Auswertungen sind aber zwei andere Limitationen ebenfalls relevant. Erstens haben wir alle Visits eines Visitors aus dem gesamten Datenbestand zusammengenommen, um die Zugehörigkeit zu einer Gruppe zu eruieren. Das verkennt, dass die gleichen Visitors über die Zeit hinweg die Gruppe wechseln könnten. So ist durchaus denkbar, dass sich im Laufe von zwei Jahren eine Benutzungspräferenz von Hauptsuchen weg und hin zu erweiterten Suchen entwickelt, z.B. weil ausgefeiltere Suchen erlernt werden. Auch ist denkbar, dass Visitors ganz unterschiedlichen Gruppen zugeordnet sind, je nachdem welches Benutzungsziel sie gerade aktuell haben (beispielsweise könnte ein*e Hochschul-lehrer*in einmal einen Text mit Wörterbuchunterstützung schreiben und ein anderes Mal Beispiele für eine Klausur suchen). Der zweite Punkt ist u.E. noch tiefgreifender, und wir können nicht oft genug darauf hinweisen: Das Konzept des Visitors ist ein rein technisches, d.h. es wird versucht, auf technischem Wege die Kluft zwischen einem angenommenen Individuum (dem Visitor) und einem tatsächlichen (ein konkreter Mensch) so klein wie möglich zu halten. Trotzdem gilt: Ein Visitor ist kein Mensch. Sogar das Mengenverhältnis ist unklar: Ein Visitor kann mehrere Individuen umfassen (z.B. wenn ein Computer von mehreren Personen genutzt wird), ein Individuum kann aber auch mehrere Visitors „generieren“, bspw. wenn es mehrere Geräte zum Zugriff nutzt oder den Browser-Cache löscht.

4 Fazit

Was bedeutet das für die eingangs definierten Fragestellungen und den Einsatz von Webanalyse-Plattformen für die Nutzungsforschung zu linguistischen Online-Ressourcen?

4.1 Externe und interne Perspektiven auf Nutzungsdaten

In der dritten Fragestellung („Was bedeutet das [= Potentiale und Limitationen] einerseits aus Sicht derer, die für den Betrieb und die (Weiter)Entwicklung von Ressourcen zuständig sind, und andererseits für externe Perspektiven auf diese Nutzungsdaten?“) erwähnen wir externe Perspektiven (also im Plural). Hier möchten wir zusammenfassend nochmals auf Wissenschaftskommunikation eingehen, dann aber auch eine neue „externe Anwendung“ von Nutzungsdaten ins Spiel bringen, nämlich institutionelle Evaluationen.

Wir haben im vorherigen Abschnitt dargelegt, welche Potentiale und Limitationen wir für die Wirksamkeitsanalyse von Wissenschaftsvermittlung in eine breitere Öffentlichkeit sehen. Zusammenfassend: Anhand der bei der Nutzung anfallenden Daten lassen sich Effekte von Disseminierungsmaßnahmen erkennen. Gleichzeitig müssen wir uns aber bewusst sein, dass dies stets ein „Indizienprozess“ ist – kausale Ketten wie „Maßnahme A führt zu Effekt B“ sind nicht ablesbar.

Die zweite Außenperspektive ist jene der institutionellen Evaluation von linguistischen Webangeboten. Dabei ist es ein berechtigtes Anliegen, dass die Entwickler*innen der (oft aus Steuergeldern) erarbeiteten linguistischen Ressourcen etwas dafür tun sollen, dass sie die jeweils adressierten Nutzer*innen auch erreichen. Doch scheinen uns Zugriffszahlen („Klickzahlen“, hier am ehesten mit Actions gleichzusetzen) gerade für wissenschaftliche Ressourcen das falsche Mittel zu sein. Zugriffszahlen üben natürlich eine gewisse Anziehungskraft aus: Sie sind einfach zu ermitteln, sie suggerieren Objektivität, und sie scheinen externes Interesse widerzuspiegeln. Dabei werden die dargestellten Limitationen unserem Eindruck nach oft nicht hinreichend beachtet. Das Design einer Ressource hat Einfluss darauf, wie viele „Klicks“ sie generiert. Wenn wir abwägen, ob eine Ressource benutzungsfreundlich gestaltet werden sollte oder ob sie möglichst viele „Klicks“ generiert, können und werden wir uns stets für die Benutzungsfreundlichkeit entscheiden, da wir – anders als kommerzielle Anbieter – uns nicht über eingeblendete Werbung finanzieren müssen. Das bedeutet, dass einfache Zugriffszahlen sowohl im Zeitverlauf zu einer Ressource, als auch im Vergleich unterschiedlicher Ressourcen nur eine sehr begrenzte Aussagekraft haben, wenn

die Ressourcen sich während der Zählung inhaltlich und technisch weiterentwickelt haben.

Grundsätzlich sollte vor dem Hintergrund der neuen Möglichkeiten digitaler Ressourcen u.E. auch ganz grundsätzlich diskutiert werden, ob Quantität von Zugriffen in der Wissenschaft tatsächlich ein sinnvolles Messinstrument ist. Wenn mit einer einschlägigen Online-Ressource einige Kolleg*innen erreicht werden, die für den Forschungszweig maßgeblich sind, kann dies wichtiger sein, als eine größere Masse von Menschen zu erreichen, die diese Informationsquelle zwar kurz, aber nicht regelmäßig benutzen. Mit anderen Worten: Orientiert man sich zu stark an den reinen Zugriffszahlen, könnte auch in wissenschaftlichen Projekten zu viel Energie z.B. in die Suchmaschinenoptimierung gesteckt werden und zu wenig in die wissenschaftliche Qualität und die gezielte Integration in die relevanten Bildungskontexte. Wir möchten dafür plädieren, dass der Nachteil solcher sog. „proxy economics“ auch für linguistische Online-Ressourcen frühzeitig diskutiert wird. Auf der einen Seite ist der Wettbewerb wichtig zur Leistungsmotivation, gleichzeitig muss das primäre Ziel die Erfüllung des gesellschaftlichen Zwecks sein (im Falle wissenschaftlicher linguistischer Online-Ressourcen: fundierte Daten zur Verfügung zu stellen, die möglichst hilfreich in den anvisierten Benutzungssituationen zu verwenden sind). Es kann nicht primär um den Proxy (hier die Zugriffszahlen) gehen, denn dadurch wird notwendigerweise das eigentliche Ziel simplifiziert. Das kommt in folgendem Zitat zum Ausdruck:

“In any ‘proxy based competition in a societal system’ an information gradient arises invariably through the measurement process. The societal goal is, by definition, what we want to achieve as a society, so it contains all the information. We might say the societal goal of the healthcare system is to maximize patient health. Now, defining what this really means in detail is extremely difficult, and may be the ever changing result of an ongoing societal discussion. The important thing to realize is that whatever it may mean, when we try to measure it we will a) lose some information and b) measure some things which we didn’t intend to measure.” (Braganza 2019)

Wir möchten nur ein paar wenige Implikationen aufzeigen, die das für die Evaluation von linguistischen Online-Ressourcen anhand von Nutzungsmetriken u.E. hat:

- Zwei verschiedene Ressourcen (bzw. Teilressourcen mit unterschiedlichem Aufbau) können in den seltensten Fällen sinnvoll miteinander verglichen werden. Besonders gilt das für die Ebene der Actions, deren Anzahl hochgradig abhängig von der Seitenarchitektur ist.
- Für verschiedene Ressourcen sollten verschiedene, auf die jeweilige Ressource zugeschnittene Metriken, gewählt werden. Für OWID berichten wir in institutionellen Evaluationen zukünftig bspw. keine Actions, sondern Visits

und Visitors. Für andere Ressourcen wären bspw. die Anzahl aktiver registrierter Benutzer*innen in einem definierten Zeitraum ein berichtenswertes Datum.¹⁶ Es sollte also der Raum gegeben werden, verschiedene Maßzahlen für Ressourcen auszuwählen und diese Auswahl inhaltlich zu begründen.

- Vergleiche über die Zeit hinweg innerhalb einer Ressource sind möglich und sinnvoll, wenn inhaltliche und technische Weiterentwicklungen zur Interpretation der Nutzungszahlen dabei hinreichend kommentiert werden können. Dabei sollte beachtet werden, dass es – gegeben die hier präsentierten Analysen – keine Überraschung ist, wenn singuläre Schwankungen den generellen Trend überdecken.

Die interne Perspektive auf Nutzungsdaten wird hauptsächlich dazu eingesetzt, linguistische Ressourcen zu einem besseren Werkzeug zur Lösung sprachlicher Probleme zu machen. Zusammenfassend kann hier konstatiert werden, dass – obwohl viele Daten aus diversen Gründen wegfallen und einige konzeptionelle Probleme auf der Hand liegen – die bei der Nutzung anfallenden Daten auf jeden Fall hypothesengenerierend eingesetzt werden können. Wir plädieren also dafür, die Ergebnisse dieser Analysen nicht für sich stehen zu lassen, sondern sie als Ausgangspunkt für weitere Untersuchungen zu nehmen. Hypothesen, die sich aus den Analysen ergeben, können einerseits als Prognosen für folgende Jahre eingesetzt werden. Andererseits können die Hypothesen aber auch dazu dienen, weitere Studien anzuschließen, in denen nicht anfallende Daten ausgewertet werden, sondern gezielt Daten *erhoben* werden.

4.2 Erhobene Daten

Anfallende Daten wie die bisher in diesem Beitrag ausgewerteten Nutzungsstatistiken sind Verhaltensspuren und sie existieren unabhängig davon, ob wir sie auswerten oder nicht. OWID wird auch benutzt, wenn wir Matomo nicht einsetzen würden, um das Verhalten der Nutzer*innen zu verfolgen. Welche Einschränkungen mit solchen Daten verbunden sein können, haben wir dargelegt. Doch wir möchten auch skizzieren, welche Alternativen wir für die Benutzungsforschung zu linguistischen Online-Ressourcen sehen. Erhobene Daten existieren nur, weil

¹⁶ Für OWID ist keine Registrierung mit Benutzernamen notwendig. Für Ressourcen mit einer solchen Registrierung ist davon auszugehen, dass die Anzahl registrierter Benutzer*innen deutlich näher an der Anzahl „echter Menschen“ liegt als das für die Visitors der Fall ist. Ein entsprechender Vergleich (Verhältnis zwischen Visitors und zugreifenden Nutzer*innen) wäre u.E. interessant für die weitere Forschung in diesem Bereich.

Forschende sich dazu entschieden haben, diese Daten zu erheben. Als Hybrid zwischen anfallenden und erhobenen Daten könnte man eine A-B-Testung sehen: Hier wird Visitors, die eine Webseite aufrufen, zufällig eine von zwei Versionen (A oder B) der Webseite angezeigt. Die Unterschiede zwischen Version A und B werden dabei typischerweise minimal gehalten, um eventuelle Änderungen im Benutzungsverhalten klar auf eine bestimmte Variation zurückführen zu können. Diese Variation kann die Größe, Einfärbung oder Positionierung einer Schaltfläche sein oder das Vertauschen von Elementen bspw. von Angabeklassen in einem Wortartikel. Dieser Teil der A-B-Testung ist somit hoch kontrolliert. Ein anderer Aspekt (und deshalb bezeichnen wir es als Hybrid) bleibt jedoch unkontrolliert, nämlich die Motivation bzw. das Benutzungsziel, mit dem die Nutzer*innen die Webseite aufrufen. Auch dieses Nutzungsziel, mit dem Besucher*innen die Webseite aufsuchen, kann einen Einfluss auf das Verhalten haben und ist insofern ein Störsignal, das wir als Forschende nicht kontrollieren können.

Im klassischen Experiment wird versucht, auch dieses Störsignal weiter einzuschränken bzw. zu eliminieren. Typischerweise müssen hier (meist persönlich präsent) Versuchsteilnehmende eine konkrete Aufgabe bearbeiten, die zuvor von den Forschenden so definiert wurde, dass die jeweilige Forschungsfrage beantwortet werden kann. Diese Aufgabe kann ebenfalls variieren, doch diese Variation ist dann Ziel des experimentellen Designs, des Versuchsplans. Auch im Experiment muss noch mit einer Vielzahl von Störvariablen (Müdigkeit, Motivation, Vorwissen usw.) umgegangen werden, doch das Ziel ist es, so viele dieser Einflüsse wie möglich zu minimieren, zu eliminieren oder zumindest zufällig über Versuchsbedingungen und -personen zu verteilen. So lassen sich, wenn darüber hinaus bestimmte Prinzipien bei der Stichprobenziehung beachtet werden, kausale Aussagen über die Wirkung bestimmter Variablen ableiten. Interessieren wir uns bspw. für die Effektivität einer neuen Suchfunktion, die evtl. eine alte Funktion ablösen soll, können wir zufällig Menschen einer von zwei Versuchsbedingungen (alte Suche vs. neue Suche) zuordnen und ihnen eine bestimmte Aufgabe geben. Wenn die Gruppe mit der neuen Suche häufiger und schneller zum korrekten Ergebnis kommt, können wir mit großer Sicherheit ableiten, dass die neue Suchfunktion der alten überlegen ist. Studien, die dieser Logik folgen, sind im Bereich der Wörterbuch-Benutzungsforschung inzwischen keine Seltenheit mehr (vgl. bspw. Dziemianko 2019; Ma & Hyeok 2018; Müller-Spitzer et al. 2014; Wolfer et al. 2016). Sie sind natürlich aufwändiger als die Aufzeichnung von Nutzungsverhalten über Webanalysetools. Wir werden für den Aufwand aber mit Ergebnissen belohnt, die hilfreicher bei Designentscheidungen sind und damit letztlich den Nutzer*innen des Webangebots zu Gute kommen können.

Literatur

- Braganza, Oliver. 2019. *Proxycconomics*. (<http://www.proxycconomics.de>) (Accessed 2020-8-6)
- Cooper, Alan, Robert Reimann & Dave Cronin. 2007. *About Face 3: The Essentials of Interaction Design*. Indianapolis, IN: Wiley.
- de Schryver, Gilles-Maurice. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16 (2). 143–199.
- de Schryver, Gilles-Maurice, David Joffe, Pitta Joffe & Sarah Hillewaert. 2006. Do dictionary users really look up frequent words?—on the overestimation of the value of corpus-based lexicography. *Lexikos* 16. 67–83.
- de Schryver, Gilles-Maurice, Sascha Wolfer & Robert Lew. 2019. The Relationship between Dictionary Look-up Frequency and Corpus Frequency Revisited: A Log-File Analysis of a Decade of User Interaction with a Swahili-English Dictionary. *GEMA Online Journal of Language Studies* 19 (4). DOI: <https://doi.org/10.17576/gema-2019-1904-01>
- Dziemiątko, Anna. 2019. An insight into the visual presentation of signposts in English learners' dictionaries online. *International Journal of Lexicography* 32 (4). 432–457. DOI: <https://doi.org/10.1093/ijl/ecv040>
- Engelberg, Stefan & Carolin Müller-Spitzer. 2013. Dictionary Portals. In Rufus H. Gouws, Ulrich Heid, Wolfgang Schweickhard, & Herbert Ernst Wiegand (eds.), *Dictionaries. An international encyclopedia of lexicography. Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*, 1023–1035. Berlin/New York: de Gruyter.
- Gillespie, Tarleton. 2010. The politics of 'platforms'. *New Media & Society* 12 (3). 347–364.
- Klosa, Annette & Rufus H. Gouws. 2015. Outer features in e-dictionaries. *Lexicographica* 31 (1). 142–172.
- Kosem, Iztok, Robert Lew, Carolin Müller-Spitzer, Maria Ribeiro Silveira, Sascha Wolfer, Amelie Dorn, ... Hilary Nesi. 2018. The Image of the Monolingual Dictionary Across Europe. Results of the European Survey of Dictionary use and Culture. *International Journal of Lexicography* 32 (1). 92–114. DOI: <https://doi.org/10.1093/ijl/ecy022>
- Lorentzen, Henrik & Liisa Theilgaard. 2012. Online dictionaries – how do users find them and what do they do once they have? In Ruth Vatvedt Fjeld & Julie Matilde Torjusen (eds.), *Proceedings of the 15th EURALEX International Congress 2012, Oslo, Norway, 7–11 August 2012*, 654–660. Oslo: Universitetet i Oslo, Institutt for lingvistiske og nordiske studier. Retrieved from <http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202012/pp654-660%20Lorentzen%20and%20Theilgaard.pdf>
- Ma, Jee Hyun & Jin Cheon Hyeok. 2018. An Experimental Study of Dictionary use on Vocabulary Learning and Reading Comprehension in Different Task Conditions. *International Journal of Lexicography* 31 (1). 29–52. DOI: <https://doi.org/10.1093/ijl/ecw037>
- Müller-Spitzer, Carolin (ed.). 2014. *Using Online Dictionaries*. Berlin, New York: de Gruyter.
- Müller-Spitzer, Carolin, Frank Michaelis & Alexander Koplenig. 2014. Evaluation of a new web design for the dictionary portal OWID An attempt at using eye-tracking technology. In Carolin Müller-Spitzer (ed.), *Using Online Dictionaries*, 207–228. Berlin, Boston: de Gruyter.
- Müller-Spitzer, Carolin, Sascha Wolfer & Alexander Koplenig. 2015. Observing Online Dictionary Users: Studies Using Wiktionary Log Files. *International Journal of Lexicography* 28 (1). 1–26. DOI: <https://doi.org/10.1093/ijl/ecu029>

- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Ripp, Saskia & Stefan Falke. 2018. Analyzing user behavior with Matomo in the online information system Grammis. In Jaka Čibej, Vojko Gorjanc, Iztok Kosem, & Simon Krek (eds.), *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*, 87–100. Ljubljana, Slovenia: Ljubljana University Press, Faculty of Arts.
- Tiberius, Carole & Jan Niestadt. 2015. Dictionary Use: A Case Study of the ANW Dictionary. *OPAL – Online Publierte Arbeiten Zur Linguistik (Research into Dictionary Use / Wörterbuchbenutzungsforschung)* 2. 28–35.
- Verlinde, Serge, Patrick Leroyer & Jean Binon. 2010. Search and you will find. From stand-alone lexicographic tools to user driven task and problem-oriented multifunctional leximats. *International Journal of Lexicography* 23 (1). 1–17.
- Wolfer, Sascha. 2020. Interessant ist, was relevant ist: Weihnachten im Wörterbuch. In Konstanze Marx (ed.), *Weihnachtslinguistik. Festliche Texte über Sprache*, 67–74. Tübingen: Narr.
- Wolfer, Sascha, Thomas Bartz, Tassja Weber, Andrea Abel, Christian M. Meyer, Carolin Müller-Spitzer & Angelika Storrer. 2016. The effectiveness of lexicographic tools for optimising written L1-texts. *International Journal of Lexicography* ecw038. DOI: <https://doi.org/10.1093/ijl/ecw038>