

## EINE DAF/Z-LERNER\*IN FRAGT NACH ...? Wie man mit empirischen Ressourcen Fragen nach undokumentierten Phänomenen beantworten kann. Ein Rundgang durch verschiedene korpusbasierte Ressourcen

Jan Oliver Rüdiger  
Leibniz-Institut für Deutsche Sprache

### Abstract

Dieser Werkstattbericht zeigt anhand verschiedener korpusbasierter Ressourcen, wie Fragen zu sprachlichen Phänomenen, die für Sprachlernende nicht oder nur unzureichend dokumentiert sind, empirisch beantwortet werden können. Besonderes Augenmerk wird dabei auf OWIDplusLIVE gelegt. Hierbei handelt es sich um ein Werkzeug zur tagesaktuellen Analyse von Token (einzelne Wortformen/Lemmata) und Bi-/Trigrammen (zwei bzw. drei direkt aufeinander folgende Token). Über eine Anbindung an KorAP können zudem Belege aus dem DeReKo (*Deutsches Referenzkorpus*) abgerufen und analysiert werden.

**Keywords:** Praxisbericht; korpusbasierte Ressourcen; empirische Analyse; Frequenzverlauf; Visualisierung

### Abstract

This hands-on article uses various corpus-based resources to show how questions about linguistic phenomena that are undocumented or poorly documented for language learners can be answered empirically. Special attention is paid to OWIDplusLIVE. This is a tool for the daily analysis of tokens (single word forms/lemmas) and bi-/trigrams (two or three directly consecutive tokens). Via a connection to KorAP, samples from the DeReKo (*German Reference Corpus*) can also be retrieved and analyzed.

**Keywords:** hands-on report; corpus-based resources; empirical analysis; frequency distribution; visualization

## 1. Einleitung

Kennen Sie den folgenden Fall: Eine DaF/Z-Lerner\*in kommt nach dem Unterricht auf Sie zu und hat eine Frage zu einem besonderen sprachlichen Phänomen, etwa zu einem der folgenden drei Fälle:

1. Welche Komposita im Deutschen mit einem bestimmten Erst- oder Zweitglied können gebildet werden? Wie unterscheiden sich diese konkreten Wortbildungen voneinander?
2. Welche Adjektiv-Nomen Kombinationen sind möglich? Welche Abfolge ist (un-)üblich?
3. Gibt es Variationen einer gegebenen Phrase? Welche Bedeutungsunterschiede sind mit der Varianz verbunden?

Bei allen drei Fällen handelt es sich zudem um sprachliche Phänomene, die nicht in einem Lehrbuch verzeichnet sind, bzw. zu denen es kein fertiges Unterrichtsmaterial gibt. Die Ursachen für fehlende Dokumentation in Lehrwerken, Wörterbüchern oder Unterrichtsmaterialien können vielfältig sein. So kann z.B. das Phänomen recht neu sein (z.B. Neologismen); es kann sich um ein Phänomen einer bestimmten Peer-Gruppen handeln (z.B. Jugendsprache – ggf. tradiert über Popkultur wie z.B. Deut-schrap); oder es handelt sich um ein komplexes, aber sehr seltenes Phänomen (ggf. auch einen Fehler,

den z.B. nur Muttersprachler\*innen machen – also nicht Teil der Sprachnorm ist), so dass dieses in der Regel nicht in Materialien für Sprachlernende dokumentiert wird.

Im Rahmen dieses Artikels sollen verschiedene korpusbasierte Ressourcen vorgestellt werden, die eine Beantwortung der oben gestellten Fälle/Fragen ermöglichen<sup>1</sup>. Das Gros dieser Ressourcen wurde zum Ziel der sprachwissenschaftlichen Forschung konzipiert – die selektive Auswahl und die folgende Anleitung ermöglichen es aber, diese Ressourcen für den eigenen Lehreinsetz zu nutzen. Je nach Niveaustufe der Lerner\*innen, den technischen Voraussetzungen und verfügbaren Ressourcen (z.B. Computer aber auch Zeit) können entweder die Lerner\*innen selbst mit diesen Ressourcen arbeiten oder die Lehrpersonen assistieren bzw. nutzen die Ressourcen, um Lernmaterial vorzubereiten.

## 2. Ein Überblick

Der Rundgang durch die verschiedenen korpusbasierten Ressourcen ist so gestaltet, dass die drei skizzierten Fälle als Grundstruktur dienen. Dabei gibt es zwei unterschiedliche Beispieltypen: Die Hauptbeispiele erklären die Grundlagen, an ihnen wird die Bedienung, die Analyse und die Ergebnisbewertung einmalig Schritt-für-Schritt durchgespielt. Zudem gibt es Illustrationsbeispiele, die weitere Verwendungsmöglichkeiten aufzeigen und zum Ausprobieren anregen sollen.

Methodische Grundlagen und Termini werden wo nötig kurz angerissen und allgemeinverständlich erklärt. Ziel ist die Nutzung der vorgestellten Ressourcen im DaF/Z-Unterricht. Für eine tiefergehende Beschäftigung werden die Einführungswerke Bubenhofer (2009), Lemnitzer / Zinsmeister (2015) und Perkuhn / Keibel / Kupietz (2012) empfohlen. Jede Ressource greift auf eine empirische Grundlage, ein Korpus zurück (Singular: Korpus, das / Plural: Korpora, die). Ein Korpus ist nach Lemnitzer / Zinsmeister (2015: 13) „[...] eine Sammlung [authentischer] schriftlicher oder gesprochener Äußerungen in einer oder mehreren Sprachen“. Für die Korpuserstellung können verschiedene Kriterien, je nach Forschungsinteresse herangezogen werden – dieser Aspekt wird aber im Folgenden ausgeblendet – zur Vertiefung sei (A) auf die oben empfohlenen Einführungswerke sowie (B) die Dokumentationen der jeweiligen Ressourcen verwiesen. In der Korpuslinguistik ist ein Korpus in der Regel immer ein digitales (d.h. maschinenlesbares) Korpus, das automatisiert ausgewertet wird. Die folgenden Ressourcen basieren also immer auf einem Korpus und sie erlauben es, über eine Web-Oberfläche mit diesem Korpus zu interagieren, es abzufragen und somit zu untersuchen.

Als erstes kurzes Beispiel soll hier das DWDS<sup>2</sup> (DWDS 2023) aufgeführt werden (siehe hierzu auch Flinz 2021). Als Online-Plattform bündelt es insbesondere historische Wörterbücher und Ressourcen. Es werden aber auch neuere Korpusquellen, wie Blogs, Webseiten oder die deutschsprachige Wikipedia integriert. Direkt über die Startseite [www.dwds.de](http://www.dwds.de) kann nach einzelnen Begriffen gesucht werden (siehe Abbildung 1, Punkt 1). Hier z.B. nach ‚Checker‘ – einem jugendsprachlichen Begriff:

---

<sup>1</sup> Der Fokus liegt dabei auf einer Einführung in die jeweiligen Ressourcen und einem Aufzeigen von Nutzungsmöglichkeiten. Die drei Fälle werden mit verschiedenen Situationen illustriert und an einem konkreten Beispiel erarbeitet.

<sup>2</sup> Slogan des DWDS: „Der deutsche Wortschatz von 1600 bis heute“.

The screenshot shows the DWDS entry for 'Checker'. At the top, there's a navigation bar with 'D W D S' and 'Der deutsche Wortschatz von 1600 bis heute.' A search bar contains 'Checker' (1). Below the search bar, the word 'Checker, der' is displayed. The entry includes grammatical information (Substantiv, Maskulinum), word breakdown (Che-cker) (2), and word origin (checken + -er) (2). The 'Bedeutung' section is under 'Technik' with 'Kontrolleur'. The 'Thesaurus' section (4) lists synonyms like 'Schlaukopf', 'Schlitzohr', 'Ausgeschlafener', 'Durchblicker', 'Fuchs', 'Schlauberger', 'Schlauerle', 'Schlaufuchs', and 'Schlaumeier'. The 'Verwendungsbeispiele für ›Checker‹' section (5) shows three authentic examples from news articles. On the right, there's a 'Worthäufigkeit' bar chart, a 'Wortverlaufskurve' line graph showing frequency from 1600 to 2000 (3), and lists of 'Weitere Wörterbücher' and 'Belege in Korpora'.

Abbildung 1  
DWDS-Eintrag zu ‚Checker‘ mit manuellen Hervorhebungen

Neben den für Wörterbüchern üblichen Verwendungs- und Bedeutungsangaben (siehe Abbildung 1, Punkt 2) erhalten Nutzer\*innen des DWDS auch Informationen zu Worthäufigkeiten (siehe Abbildung 1, Punkt 3 Wortverlaufskurve, die eine klare Zunahme der Verwendungshäufigkeit nahelegt), Daten aus verknüpften Ressourcen (siehe Abbildung 1, Punkt 4 – z.B. Synonyme aus dem verknüpften OpenThesaurus<sup>3</sup>) und Korpusbelege (siehe Abbildung 1, Punkt 5 – Verwendungsbeispiele). Bei den Verwendungsbeispielen handelt es sich um reale und damit authentische Beispiele für die Verwendung. Gerade für einen Einstieg in die Arbeit mit korpusbasierten Ressourcen eignet sich das DWDS sehr gut, da sowohl die Bedienung als auch Auswertung sehr nah an einem regulären Wörterbuch orientiert sind.

Eine ähnliche, aber stärker an der Gegenwartssprache orientierte Wörterbuch-Plattform bietet OWID (OWID 2023). Mit [www.OWID.de](http://www.OWID.de) wird eine Suche (siehe Abbildung 2, Punkt 1: Suchbegriff

<sup>3</sup> [www.OpenThesaurus.de](http://www.OpenThesaurus.de) ist ein gemeinschaftlich erarbeiteter und öffentlicher Thesaurus – der genauso kollaborativ entsteht wie die Wikipedia.

,Weg‘ / 2: Suchergebnisse) über verschiedene sprachwissenschaftliche Wörterbücher (siehe Abbildung 2, Punkt 3: Liste aller Ressourcen) möglich. Außerdem bietet OWID mit OWIDplus (siehe Abbildung 2, Punkt 4) einen gesonderten Bereich, in dem sich experimentelle Sprachressourcen finden lassen.

**OWID**  1

IDS

**Ergebnis für 'Weg'** 2

Siehe Artikel  
| [weg](#) (elexiko)  
| [Weg](#) (elexiko)  
Weg Basiskomponente zu [Der Weg zur Hölle ist mit guten Vorsätzen gepflastert.](#) (Sprichwörterbuch)  
Weg Basiskomponente zu [Der Weg ist das Ziel.](#) (Sprichwörterbuch)  
Weg Basiskomponente zu [Wo ein Wille ist, ist auch ein Weg.](#) (Sprichwörterbuch)  
Weg Basiskomponente zu [Wir machen den Weg frei.](#) (Sprichwörterbuch)  
Weg Basiskomponente zu [Die Wege des Herrn sind unergründlich.](#) (Sprichwörterbuch)  
Weg Basiskomponente zu [Alle Wege führen nach Rom.](#) (Sprichwörterbuch)

Stichwörter, die mit 'Weg' anfangen:  
| Schlüsselwörter der Wendezeit (7) ▾  
| [Weggang](#) (Schlüsselwörter der Wendezeit)  
| [weggehen](#) (Schlüsselwörter der Wendezeit)  
| [weglaufen](#) (Schlüsselwörter der Wendezeit)  
| [wegmachen](#) (Schlüsselwörter der Wendezeit)  
| [wegrennen](#) (Schlüsselwörter der Wendezeit)  
| [wegziehen](#) (Schlüsselwörter der Wendezeit)  
| [Wegzug](#) (Schlüsselwörter der Wendezeit)  
| Verlaufsformenwörterbuch (7) ▶  
| elexiko (400) ▶  
| Feste Wortverbindungen (1) ▾  
Wegbereiter Basiskomponente zu [geistiger Wegbereiter](#) (Feste Wortverbindungen)  
| Neologismenwörterbuch (26) ▶

Stichwörter, die auf 'Weg' enden:  
| elexiko (233) ▶  
| Neologismenwörterbuch (11) ▾  
[Barfußweg](#) siehe 'Sprachreflexives' in [Barfußpfad](#) (Neologismenwörterbuch)  
[Barfußweg](#) siehe 'Sinverwandte Ausdrücke' in [Barfußpfad](#) (Neologismenwörterbuch)  
[Baumkronenweg](#) siehe 'Sprachreflexives' in [Baumwipfeldpfad](#) (Neologismenwörterbuch)  
[Baumkronenweg](#) siehe 'Sinverwandte Ausdrücke' in [Baumwipfeldpfad](#) (Neologismenwörterbuch)  
[Waldwipfelweg](#) siehe 'Sprachreflexives' in [Baumwipfeldpfad](#) (Neologismenwörterbuch)  
[Waldwipfelweg](#) siehe 'Sinverwandte Ausdrücke' in [Baumwipfeldpfad](#) (Neologismenwörterbuch)  
drückte weg 1./3. Person Singular Indikativ Präteritum von [wegdrücken](#) (Neologismenwörterbuch)  
Heilwaldweg siehe 'Wortbildungsproduktivität' in [Heilwald](#) (Neologismenwörterbuch)  
klickte weg 1./3. Person Singular Indikativ Präteritum von [wegklicken](#) (Neologismenwörterbuch)  
[Radschnellweg](#) (Neologismenwörterbuch)  
zappte (sich) weg 1./3. Person Singular Indikativ Präteritum von [wegzappen](#) (Neologismenwörterbuch)

3  
4  
| elexiko  
| Paronymwörterbuch  
| Sprichwörterbuch  
| Kommunikationsverben  
| Verlaufsformen  
| Fremdwörterbuch  
| Neologismenwörterbuch  
| Demokratiediskurs 1918-25  
| Schuldiskurs 1945-55  
| Protestdiskurs 1967/68  
| Schlüsselwörter 1989/90  
| OBELEX<sup>meta</sup>  
| OBELEX<sup>dict</sup>  
| Korpusuche  
| OWID<sup>plus</sup>

Abbildung 2  
OWID-Suche zu ‚Weg‘ mit manuellen Hervorhebungen

Jede Ressource auf OWIDplus wird mit einer kurzen Einführung vorgestellt. Die Ressourcen bieten dann je nach Konzeption weiterführende Informationen an. Unterhalb des Symbolbilds für die Ressource ist der Ressourcennamen vermerkt (siehe Abbildung 3, Punkt 1). Ein kurzer Beschreibungstext erklärt, das grobe Konzept und mögliche Verwendungsweisen (siehe Abbildung 3, Punkt 2). Jede Ressource ist mit einer Liste von Schlagworten verknüpft (siehe Abbildung 3, Punkt 3). Auf diese kann geklickt werden, um ähnliche Ressourcen zu finden. Eine komplette Liste der Schlagworte kann mit einem Klick auf ‚Filter‘ (siehe Abbildung 3, Punkt 4) angezeigt werden. Über den entsprechenden Button (siehe Abbildung 4, Punkt 5) gelangt man dann zur jeweiligen Ressource.



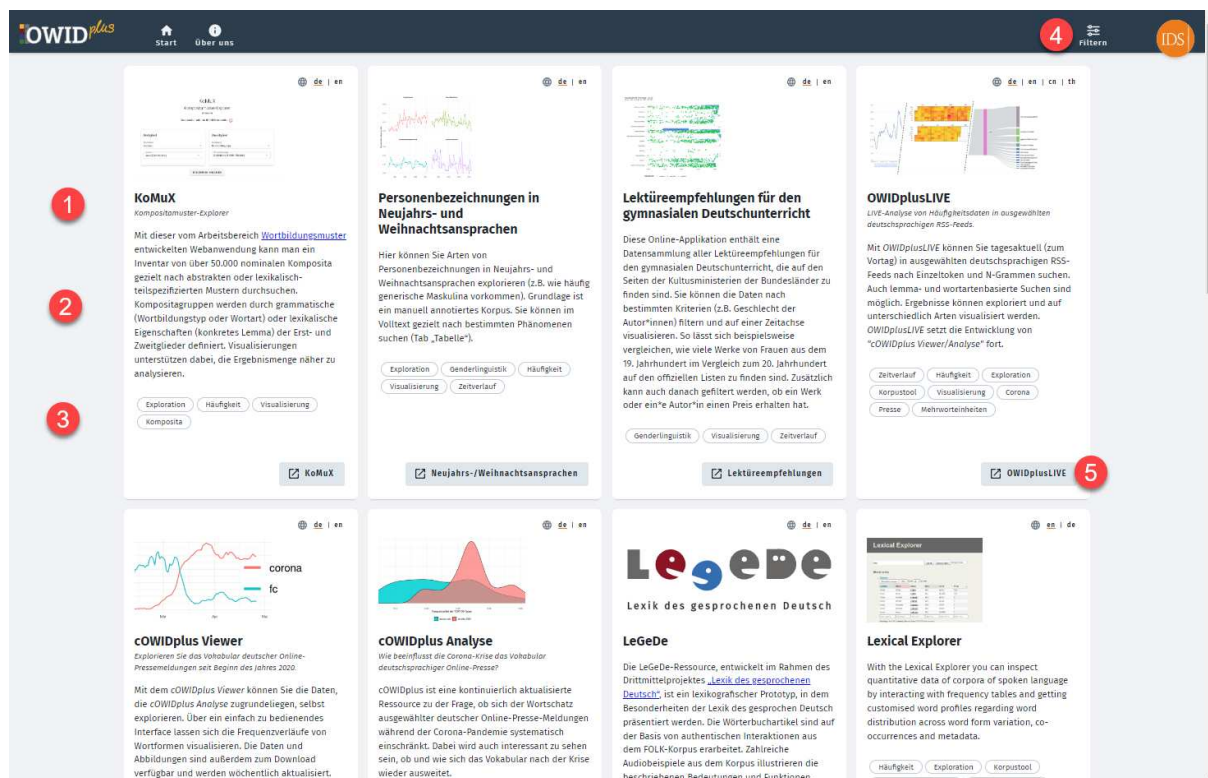


Abbildung 3  
OWIDplus mit manuellen Hervorhebungen

Besonders reizvoll für den Unterrichtseinsatz von OWID und OWIDplus ist die große Bandbreite an verfügbaren Ressourcen (z.B. zu gesprochener und geschriebener Sprache; einzelnen Analysen; komplexe Datensammlungen zu verschiedenen Gegenstandsbereichen und Phänomenen). Für die eingangs vorgestellten Fälle/Fragen soll primär die Ressource OWIDplusLIVE genutzt werden (Aufruf über einen Klick auf den entsprechenden Button – siehe Abbildung 3, Punkt 5). Mit diesem Tool können tagesaktuelle (zum Vortag) Analysen durchgeführt werden (seit 2020). Bei ‚klassischen‘ Korpora liegen zwischen Korpuserstellung und Auswertung meist mehrere Wochen, Monate oder sogar Jahre<sup>4</sup>. Daher eignet sich OWIDplusLIVE besonders für aktuelle Phänomene. Die Web-Oberfläche ist nutzerfreundlich gestaltet und es gibt ein kurzes deutschsprachiges Einführungsvideo, das die wesentlichen Funktionen erklärt.

OWIDplusLIVE umfasst ein Korpus mit aktuell (Stand: 17.11.2023) rund 116 Mio. Token. Als Token werden hierbei alle Wortformen und Satzzeichen gezählt. Bei 400 Token pro Seite entspräche dies in etwa 252 Tausend vollständig bedruckte Seiten oder bei 300 Seiten pro Buch in etwa 840 Büchern. Grundlage für das Korpus bilden kontinuierlich erfasste RSS-Nachrichtenfeeds von 13 deutschsprachigen Zeitungen (11 Deutschland / je eine Zeitung aus Österreich und der Schweiz)<sup>5</sup>. RSS-Nachrichten sind Kurzmeldungen, die von den Zeitungen zu den einzelnen Artikeln veröffentlicht werden und frei zugänglich sind. OWIDplusLIVE sammelt die Daten fortlaufend und bündelt die Daten am Tagesende zu einem Datenpunkt. So lassen sich sprachliche Phänomene seit dem

<sup>4</sup> Erwähnenswerte Ressourcen, die aktuell nahezu tagesaktuelle Analysen bieten bzw. diese Art der Echtzeitanalyse sprachwissenschaftlich erproben: Davies (2013); Vogel et al. (2021); Barbaresi (2022).

<sup>5</sup> Für Hintergrunddetails zur Auswahl der Zeitungen, der Aufbereitung und verschiedenen Analysen siehe Wolfer (2020); Wolfer et al. (2020).

01.01.2020 in den 13 genannten Quellen untersuchen. Die Korpusdaten werden zudem automatisiert mit Lemma- und POS<sup>6</sup>-Informationen angereichert. Die erlaubt es z.B. nach der Grundform (z.B. *lieben* zu suchen, um alle Wortformen z.B. *lieben, geliebt, liebte, etc.* zu finden) oder einer Eingrenzung für eine bestimmte Wortart (z.B. Grundform: *modern* + POS: Adjektiv führt zu den Wortformen: *modern, moderne* (etwas ist *modern*) – Grundform: *modern* + POS: Verb führt zu den Wortformen: *modern, moder, modert* (etwas *modert* – im Sinne von Verwesung oder Fäulnis)).

### Fall 1: Neologismen sowie Suche nach Erst-/Zweitgliedern in Komposita

Für diesen ersten Fall sind mehrere konkrete Möglichkeiten denkbar, z.B.: Könnte die DaF/Z-Lerner\*in nach einem konkreten Neologismus fragen; oder welche Komposita es im Deutschen im Rahmen der COVID-19-Pandemie gegeben hat<sup>7</sup>; oder ob ein bestimmtes Wort eher als Erst- oder eher als Zweitglied gebräuchlich ist.

OWIDplusLIVE kann entweder über [www.owid.de](http://www.owid.de) > OWIDplus > OWIDplusLIVE oder über die direkte URL<sup>8</sup> aufgerufen werden. Wenn das erste Mal OWIDplusLIVE startet, erscheint ein Einführungsvideo<sup>9</sup> – dieses kann mit „Tutorial beenden“ geschlossen werden<sup>10</sup>. Die folgende Eingabemaske wird angezeigt:

The screenshot shows the OWIDplus search interface. At the top left is the logo 'OWID plus' and 'IDS LEIBNIZ-INSTITUT FÜR DEUTSCHE SPRACHE'. Below the logo is a search bar with the text 'SUCHEN: Einfache Suche'. Underneath the search bar are three radio buttons for window size: 'N=1' (selected), 'N=2', and 'N=3'. Below that is a dropdown menu for search type, currently set to 'Wortform'. To the right of the dropdown is a search input field containing the text 'Angstgegner'. Below the input field are several example search queries: 'Wortform = deutschland | Lemma = sagen | POS = ART | Wortform = \*sagte | Lemma = \*bayer\* | POS = KO\*'. At the bottom left is a button labeled 'ALLE EINGABEN LÖSCHEN' and at the bottom right is a button labeled 'SUCHE AUSFÜHREN'. Five red circles with numbers 1 through 5 are overlaid on the interface to highlight these features.

Abbildung 4  
Suchmaske OWIDplusLIVE mit manuellen Hervorhebungen

<sup>6</sup> POS steht hierbei für ‚Part-of-Speech‘ bzw. die Wortart wie z.B. Nomen, Adjektiv, etc. – Die POS-Annotation erfolgt mittels TAGs, d.h. für die jeweilige Wortart existiert ein TAG, ein Kürzel. Das von OWIDplusLIVE verwendete TAG-Set (Sammlung aller TAGs) ist das STTS (Anne Schiller u. a. 1999). OWIDplusLIVE bietet eine integrierte Nachschlagefunktion für die STTS-TAGs an, auf die später eingegangen wird.

<sup>7</sup> Hier sei anzumerken, dass kein Korpus eine Aussage über die Gesamtheit (also z.B. das Deutsche) geben kann (methodische Begrenzung). Selbst sehr große Korpora wie das *Deutsches Referenz Korpus* – kurz DeReKo (vgl. Kupietz 2020) erlauben keine derartigen Aussagen. Dies ist in der Regel auch nicht nötig. Auch wenn sich mit OWIDplusLIVE nur ein Teil z.B. eine Auflistung der häufigsten Corona/COVID-19 Komposita ermitteln lässt, so reicht dies in aller Regel aus, um damit produktiv zu arbeiten, zu analysieren und Unterrichtsmaterial zu erstellen.

<sup>8</sup> Siehe: <https://www.owid.de/plus/live-2021/> (11.11.2023).

<sup>9</sup> Die Betrachtung wird empfohlen.

<sup>10</sup> Falls das Video zu einem späteren Zeitpunkt (erneut) betrachtet werden soll – kann das Video mit einem Klick auf das Hauptmenü (Abbildung 9, Punkt 8) aufgerufen werden. Im Hauptmenü muss dann der Menüpunkt „Hilfe anzeigen“ ausgewählt werden.

In dieser Eingabemaske stehen folgende Funktionen zur Verfügung: Als erstes (Abbildung 4, Punkt 1) kann die Suchfenstergröße ausgewählt werden, d.h. es kann ein einzelnes Token (N=1), zwei direkt aufeinander folgende Token (N=2, ein so genanntes Bi-Gramm) und drei direkt aufeinanderfolgende Token (N=3, ein so genanntes Tri-Gramm) gesucht werden. Im konkreten Beispiel werden die Standardeinstellungen (Abbildung 4, Punkt 1: N=1 und Abbildung 4, Punkt 2: Wortform) belassen und in das Eingabefeld (Abbildung 4, Punkt 3) wird das Komposita ‚Angstgegner‘ eingetragen<sup>11</sup>. Zum Starten der Suchen erfolgt ein Klick auf „Suche ausführen“ (Abbildung 4, Punkt 5). Während die Suche läuft, erscheint eine Ladeanzeige. Nach wenigen Sekunden erscheinen dann die Ergebnisse unterhalb des Suchfelds (Abbildung 5, Punkt 1):

---

<sup>11</sup> Weitere Beispiel-Suchen sind unter (Abbildung 4, Punkt 4) zu finden.

**OWIDplus** IDS LEIBNIZ-INSTITUT FÜR DEUTSCHE SPRACHE

N=1 N=2 N=3

Suchabfrage: Suchausdruck (einzelnes Token)

Wortform  1

Beispiele (anklicken und ausführen):  
Wortform = deutschland | Lemma = sagen | POS = ART | Wortform = \*sagte | Lemma = \*bayer\* | POS = KO\*

ALLE EINGABEN LÖSCHEN SUCHE AUSFÜHREN

SUCHEN: Erweiterte Tiefen-Suche

(N = 1): [1. Wortform] = angstgegner [1 / 1]

ANPASSEN: Einstellungen für Visualisierung ändern

Relative Frequenz

Granulierung: Tag Woche Monat Quartal Jahr Glättung: 1 7 14 21 28 31

SUCHVERLAUF: Vorangegangene Suchen vergleichen & verfeinern

(N = 1): [1. Wortform] = angstgegner

	Wortform	Lemma	POS-Tag	Erfasst (Tage)	Erfasst (%-Tage)	Summe	Summe (rel.)	Frequenzkurve (rel.)	KorAP
<input checked="" type="checkbox"/>	angstgegner	angstgegner	NN	147	12.28070	274	2.70391		

Zellen pro Seite: 5 1-1 von 1

Korpusgröße (Stand: 2023-04-28):  
N=1: 101334554 / N=2: 91647255 / N=3: 82086084

2023 — OWIDplusLIVE | Kontakt | Impressum | Datenschutzhinweis

Abbildung 5  
OWIDplusLIVE - Suche nach ‚Angstgegner‘ (Stand: 28.04.2023) mit manuellen Hervorhebungen

Im Visualisierungsbereich (Abbildung 5, Punkt 2) ist ein Frequenzverlauf zu sehen. Am Ende der Seite listet der „Suchverlauf“ (Abbildung 5, Punkt 3) alle bisherigen Suchen zur Suchfenstergröße N=1 auf und die dazu passenden Übereinstimmungen.



Zum Vergleich mehrerer Tokens kann die Suche mehrfach hintereinander ausgeführt werden. Bsp.: Ein erneutes Ausführen der Suche – zu Layer: Lemma (Abbildung 4, Punkt 2) und ‚Abstiegsangst‘ (Abbildung 4, Punkt 3) führt dazu, dass im Visualisierungsbereich (Abbildung 5, Punkt 2) zwei Zeitverlaufskurven angezeigt werden. Im Suchverlauf (Abbildung 5, Punkt 3) zeigt sich zudem, dass es für das Lemma ‚Abstiegsangst‘ zwei Wortformen (Realisierungen) gibt: Absteigsangst und Absteigs-ängste<sup>12</sup>. Eine weitere Möglichkeit besteht darin, s.g. *Wildcards* (Platzhalter) zu verwenden. Dazu verwendet man einfach das Zeichen \*. So findet eine Suche nach ‚Angst\*‘ alle Wortformen, die mit ‚Angst‘ beginnen (z.B. Komposita mit Angst als Erstglied) – eine Suche nach ‚\*angst‘ alle Wortformen, die auf ‚angst‘ enden (z.B. Komposita mit Angst als Zweit-/Letztglied). Beispiel für Ergebnisse aus dem Suchverlauf für ‚\*angst‘ (Stand: 28.04.2023):

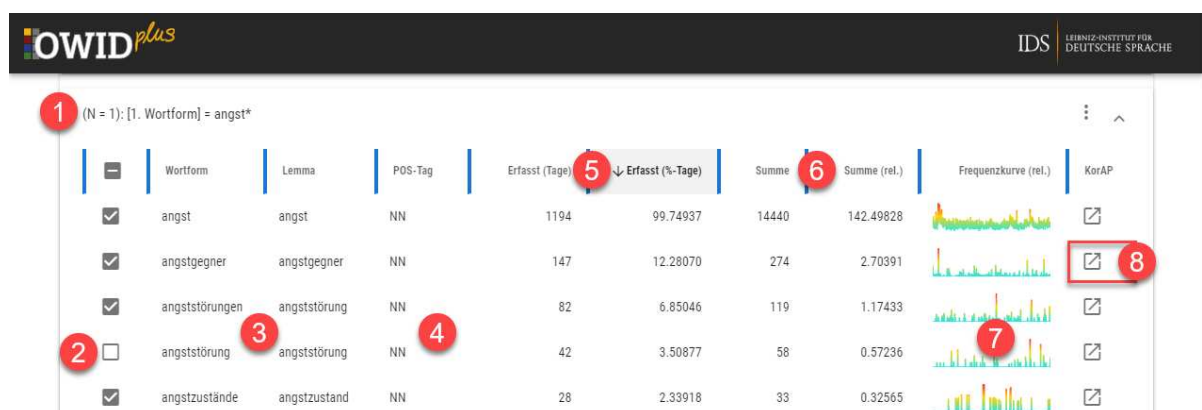


Abbildung 6  
Ausschnitt der Ergebnisse zu ‚angst\*‘ mit manuellen Hervorhebungen

In (Abbildung 6, Punkt 1) ist die Abfrage kurz zusammengefasst zu sehen. Jedes mit der Abfrage verbundene Ergebnis wird einzeln aufgelistet. Gelegentlich kommt es vor, dass eine Abfrage zu viele oder ungewünschte Ergebnisse liefert. Es ist daher möglich, einzelne Ergebnisse abzuwählen, indem man den Haken vor dem jeweiligen Ergebnis entfernt (siehe Abbildung 6, Punkt 2). Die Ergebnisse unterscheiden sich auf mindestens einem Layer (Wortform, Lemma und/oder POS), siehe hierzu (Abbildung 6, Punkt 3 und 4). Für jedes Ergebnis wird ausgewiesen, an wie vielen Tagen (auch in Prozent) der entsprechende Ergebniseintrag im Korpus enthalten ist. Da es pro Tag zu mehr als einer Übereinstimmung kommen kann, wird die Summe aller Vorkommen, sowie die relative Summe (Summe rel.) angegeben (siehe Abbildung 6, Punkt 6)<sup>13</sup>. Ein verkleinerter relativer Frequenzverlauf (siehe Abbildung 6, Punkt 7) kann Hinweise darauf geben, ob das jeweilige Ergebnis zu bestimmten Zeitpunkten (starker roter Ausschlag nach oben) einen Effekt auf die Gesamtmenge haben kann. Da OWIDplusLIVE nur eine Ressource zum Nachschlagen von Token, Bi- und Tri-Grammen ist und keine Belegstellen bietet, wurde die das Korpus-Recherchesystem KorAP angebunden (siehe Abbildung 6, Punkt 8). Mit einem Klick auf den entsprechenden Button wird zunächst eine Anmeldung an

<sup>12</sup> Hinweis: OWIDplusLIVE normalisiert alle Token – dabei werden alle Token klein geschrieben. Homographie (siehe Beispiel ‚modern/modern‘ oben) können über das POS-Tag in der Regel ausdifferenziert werden.

<sup>13</sup> Die Angabe *Summe* ist eine absolute Zahl – d.h. exakt so oft kommt der Eintrag im Korpus vor. Da die Korpusmenge pro Tag und somit zwischen unterschiedlichen Zeitabschnitten schwankt, wird auch die relative Frequenz (absolute Frequenz geteilt durch Summe aller Token pro Zeitabschnitt) angegeben. Eine Angabe in Prozent ist eine relative Zahl – der Bezugswert ist hier 100. OWIDplusLIVE gibt die relative Frequenz immer in ‚pro Mio. Token‘ an, d.h. der Bezugswert ist hier 1 Mio.

das KorAP-System<sup>14</sup> angefordert. Ist die Registrierung erfolgt, werden Belege aus DeReKo (Deutsches Referenzkorpus) abgerufen<sup>15</sup>:



Abbildung 7  
Auswahl von DeReKo Belegen zu 'Angstgegner' in OWIDplusLIVE

Die Belegsätze sind aligniert (siehe Abbildung 7, Punkt 1). Die Übereinstimmungen stehen dabei exakt untereinander, mit einer zusätzlichen eindeutigen Text-Kennung (z.B. WUD17/B82/98528)<sup>16</sup>. Durch die Alignierung, einer s.g. KWIC-Darstellung<sup>17</sup> lassen sich durch Lesen schnell die mit dem Suchwort verbundenen Kontexte erschließen bzw. Sprachgebrauchsmuster identifizieren. Im konkreten Fall wird klar, dass Angstgegner sich in der Vielzahl aller Fälle auf einen starken evtl. überlegenen Gegner vorzugsweise im Sport-Kontext handelt. Mit einem weiteren Klick auf den Button (siehe Abbildung 7, Punkt 2) kann KorAP direkt genutzt werden, um eine vertiefende und erweiterte Abfrage zu realisieren<sup>18</sup>. Mit einem Klick auf den Schließen-Button (Abbildung 7, Punkt 3) können die Belegstellen ausgeblendet werden.

Wenn also z.B. die DaF/Z-Lerner\*in nach der Bedeutung von ‚Angstgegner‘ fragt, z.B., weil dieser Begriff im Kontext einer gesehenen Fußballübertragung auf Deutsch gefallen ist, gibt es mehrere Ansatzpunkte für eine Recherche in den vorgestellten Ressourcen: (1) Das DWDS bietet einen Wörterbucheintrag zu ‚Angstgegner‘. (2) OWIDplus bietet zwar mit dem *Kicktionary* ein kleines Wörterbuch mit 2000 Einträgen zur Fußballsprache aber keinen Eintrag zum gesuchten Wort. (3) OWIDplusLIVE hat das Token an 147 Tagen (12,2% der Tage) verzeichnet (Stand: 28.04.2023) und liefert weitere Informationen zur Häufigkeit (4) KorAP liefert zudem über 31800 Belege aus einer Vielzahl an Quellen. Nach der manuellen Sichtung, also der Durchsicht und der Lektüre der Belege aus einer Stichprobe von 100 Beispielen, deutet alles darauf hin, dass es sich bei dem Token ‚Angstgegner‘ um einen Sportjargon handelt und dass die Verwendung in anderen Kontexten eher unüblich ist.

<sup>14</sup> Eine kostenlose Registrierung ist hierfür notwendig.

<sup>15</sup> Hinweis: Die Quellen (RSS-Feeds) in OWIDplusLIVE sind nicht Teil von DeReKo. Daher weichen die Ergebnismengen voneinander ab (mehr oder weniger viele Ergebnisse). Die Nutzung von KorAP/DeReKo erlaubt aber gerade die gewinnbringende Suche nach interessanten Belegbeispielen.

<sup>16</sup> Diese Text-Kennung erlaubt das Auffinden eines Beleges/Textes in DeReKo.

<sup>17</sup> KWIC steht hierbei für „Keyword in Context“. Das Keyword – bzw. die exakte Übereinstimmung steht im Satz-Kontext.

<sup>18</sup> KorAP verfügt über eine komplexe Suche mit vielen Möglichkeiten. Zur Vertiefung sei auf die in KorAP direkt integrierte Dokumentation verwiesen – siehe: KorAP - Team (2023).

## Fall 2: Welche Adjektiv-Nomen Kombinationen sind üblich (Gebrauchshäufigkeit)?

Die Abfrage von einzelnen Token in klassischen Wörterbüchern führt durchaus oft zum Erfolg<sup>19</sup>. Für Wortkombinationen gibt es zudem spezialisierte Wörterbücher (z.B. Sprichwörterbücher<sup>20</sup> oder Kollokationswörterbücher). Trotzdem kann es passieren, dass in diesen keine alltagssprachlichen Phänomene verzeichnet sind.

Im Folgenden werden zwei Fragen parallel untersucht: (A) Gibt es neben dem bekannten Bi-Gramm „verdienter Sieg“ auf eine (analog) „verdiente Niederlage“? (B) Gibt es im authentischen Sprachgebrauch<sup>21</sup> Flexionsformen von Farbadjektiven, die es nach einer normativen Grammatik nicht geben sollte – z.B. zu pink, lila oder orange (z. B. weil einige Muttersprachler dies so verwenden)?

The screenshot shows the OWIDplus search interface. At the top, it says 'OWIDplus' and 'IDS | LEIBNIZ-INSTITUT FÜR DEUTSCHE SPRACHE'. Below that, there's a search bar with 'SUCHEN: Einfache Suche'. Underneath, the search window size is set to 'N=2', with 'N=1' and 'N=3' also visible. A red circle with the number '1' is around the 'N=2' selection. Below that, the search query is shown in two slots. The first slot is labeled 'Suchausdruck (an 1. Position)...' and contains 'ADJA'. A red circle with the number '3' is around this slot. The second slot is labeled 'Suchausdruck (an 2. Position)...' and contains 'Sieg'. A red circle with the number '2' is around the 'POS' dropdown menu, which is currently set to 'ADJA'. A red arrow points from the 'POS' dropdown to the 'ADJA' text in the first slot. At the bottom, there are two buttons: 'ALLE EINGABEN LÖSCHEN' and 'SUCHE AUSFÜHREN'.

Abbildung 8

Suche nach dem Bi-Gramm „Adjektiv + Sieg“ in OWIDplusLIVE mit manuellen Hervorhebungen

Zunächst muss die Suchfenstergröße auf N=2 eingestellt werden, um nach Bi-Grammen suchen zu können (siehe Abbildung 8, Punkt 1). Durch die Umstellung erscheinen zwei Eingabeslots. Für den ersten Eingabeslot ist anstelle von ‚Wortform‘ der Layer ‚POS‘ auszuwählen (mit einem Klick auf den Pfeil nach unten, rechts neben Wortform – Abb. 8 – 2). Sobald ‚POS‘ ausgewählt wurde, erscheint oberhalb des Auswahlpfeils ein kleines Info-Symbol. Auf dieses kann geklickt werden, um eine Liste mit gültigen POS-TAGs sowie Beispielen zu erhalten. Für die Suche nach attributiven Adjektiven kann des TAG ‚ADJA‘ verwendet werden. Für den zweiten Suchslot kann dann die Wortform ‚Sieg‘ bzw. ‚Niederlage‘ eingetragen werden (vgl. Abbildung 8, Punkt 3). Auch hier ist eine nacheinander gestellte Suchabfrage möglich – z.B. zuerst nach ADJA + Sieg und dann nach ADJA + Niederlage.

Für die Abfragen ADJA + Sieg bzw. Niederlage gibt es eine Vielzahl von möglichen Kombinationen in OWIDplusLIVE. Eine Durchsicht des Suchverlaufs (vgl. Abbildung 5, Punkt 3) ergibt, dass sowohl „verdienter Sieg/Niederlage“ als auch die gesteigerte Form „hochverdiente/r Sieg/Niederlage“ belegt sind. Die Suche (Stand: 29.04.2023) zu ‚Sieg‘ liefert 616 mögliche Adjektiv-Sieg-Kombinationen, die Suche zu ‚Niederlage‘ 339 Adjektiv-Niederlage-Kombinationen. Folgende Punkte sind bei der Reflektion der Ergebnisse zu beachten:

<sup>19</sup> Siehe auch Beispiel (oben) zu ‚Angstgegner‘ im DWDS.

<sup>20</sup> Siehe z.B. Sprichwörterbuch auf [www.OWID.de](http://www.OWID.de).

<sup>21</sup> Da Korpora auf authentischen Sprachdaten basieren (s.o.) lassen sich so auch Phänomene des allgemeinen Sprachgebrauchs untersuchen.

1. Die von OWIDplusLIVE vorgeschlagene Reihenfolge orientiert sich an der Erfassungshäufigkeit. Dadurch tauchen häufige Kombinationen wie z.B. „erster/zweiter/dritter Sieg“ in den oberen Listenrängen auf. Gelegentlich muss man die ganze Liste durchsehen, um interessante Ergebnisse zu identifizieren.
2. OWIDplusLIVE ist, wie bereits erwähnt, nur ein kleines Korpus und speist sich aus lediglich 13 Quellen. D.h. auch wenn eine Kombination nur wenige Male belegt ist, kann diese Kombination häufig und regelmäßig genutzt werden. Zum Abgleich kann die KorAP-Suche verwendet werden. Oft finden sich via KorAP tausende konkrete Belegstellen in einer wesentlich größeren Daten- und Quellenbasis.
3. Evtl. finden sich sowohl in OWIDplusLIVE als auch in KorAP kleinere Datenfehler. Dies können z.B. sinnlos erscheinende Zeichenketten (z.B.: +++) sein oder die automatische Annotation der POS-TAGs ist fehlerhaft. Diese Fehler sind in einem vollautomatischen Prozess und zudem bei authentischem Sprachmaterial<sup>22</sup> unvermeidlich. Für vertiefende korpuslinguistische Analysen gibt es Verfahren zur Fehlerbereinigung, auf die hier nicht eingegangen werden muss – zur Vertiefung sind Heyer / Quasthoff / Wittig (2006), Bubenhofer (2009), Lemnitzer / Zinsmeister (2015) und Perkuhn / Keibel / Kupietz (2012) empfohlen. In einem Analyseprozess zur Konzeption von DaF/Z-Unterrichtsmaterial können diese Fehler relativ problemlos ignoriert werden.

Wie das Beispiel zeigt, können die Ergebnismengen (616 + 339) umfangreich sein. Daher ist es oft sinnvoll, eine Suchanfrage möglichst eng zu formulieren. Wenn eine Recherche nach der zweiten Frage gestartet werden soll, dann ist für den Suchslot 1: Lemma (anstelle von POS) und der Suchwert z.B. *orange*, *pink* oder *lila* einzutragen – und nicht POS = ADJA. Für Suchslot 2 wird dann die Einstellung POS und der Suchwert NN (Normales Nomen) empfohlen. Es finden sich so z.B. Einträge zu „orange Haut“, „orange Kürbis“ oder „orange Brille“ – ebenso wie verschiedene Wortformen „orangene Breireste“, „orangenes Quadrat“ oder „orangenes Messeflugzeug“ – vgl. hierzu Dudenredaktion (2009: 344-345).

Bisher wurde der Visualisierungsbereich (siehe Abbildung 5-2) ausgespart und nur mit dem Suchverlauf (Abbildung 5, Punkt 3) gearbeitet, da dieser für die Fragestellungen dieses Beitrags meist interessantere Details durch Nutzung von KorAP-Belegen bietet. Trotzdem soll ein kurzer Blick auf diesen Bereich geworfen werden, um sprachliche Phänomene zu identifizieren, die mit einem Ereignis verbunden sind (z.B. Corona-Pandemie, russischer Angriffskrieg, Sportereignis, etc.). Für das folgende Beispiel wird nach dem Bi-Gramm (N=2) 1. POS = ADJA und 2. Wortform = Weihnachten gesucht. Die häufigste Realisationsform (Stand: 30.04.2023) ist „weiße Weihnachten“ noch vor „frohe Weihnachten“. Im Visualisierungsbereich ist die folgende Kurve zu sehen:

---

<sup>22</sup> Da es sich um authentisches Sprachmaterial handelt, werden Fehler im Originaltext übernommen. Außerdem kann es während des automatischen Prozesses (Korpusaufbereitung) zu Fehlern kommen – z.B. dadurch, dass Token nicht korrekt getrennt werden.

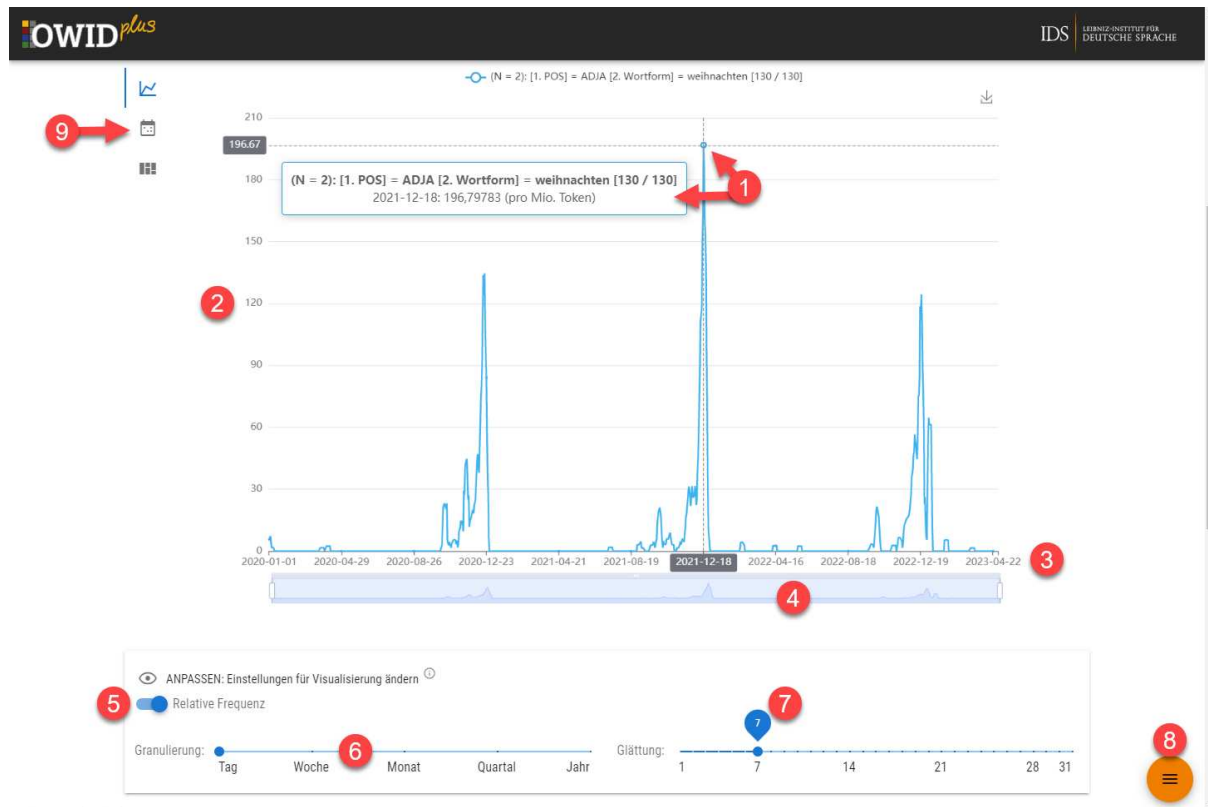


Abbildung 9

Suche zu  $N=2$  [1. POS] = ADJA / [2. Wortform] = weihnachten – mit manuellen Hervorhebungen

Die Grafik visualisiert in der Standardeinstellung alle Ergebnisse der Abfrage(n)<sup>23</sup>. Gut zu erkennen sind die frei markanten Ausschläge – die exakt mit den drei Weihnachten im bisherigen Korpus übereinstimmen. Wird der Mauscursor über die Grafik bewegt, erscheinen zu den Punkten der Kurve (unterhalb des Mauszeigers) weiterführende Informationen (siehe Abbildung 9, Punkt 1). In der Grafik wird z.B. ersichtlich, dass am 18. Dezember 2021<sup>24</sup> eine relative Frequenz<sup>25</sup> für die Suchabfrage von ca. 196,79 (pro Mio. Token) vorlag. Die vertikale Y-Achse (Abbildung 9, Punkt 2) gibt die (relative) Frequenz an, sie passt sich automatisch dem Maximalwert an. Die horizontale X-Achse (Abbildung 9, Punkt 3) ist der Zeitverlauf, beginnend links am 01.01.2020 nach rechts (Vortag). Unterhalb der X-Achse befindet sich ein Zoom-Bereich (Abbildung 9, Punkt 4), mit dem sich der Zeitabschnitt eingrenzen lässt (Ankerpunkte links und rechts können mit der Maus verschoben werden). Der Visualisierungsbereich kann zudem über die Schaltflächen unter „Anpassen“ modifiziert werden. Vorrangig der Schalter „Relative Frequenz“ (Abbildung 9, Punkt 5), der ein An- und Abschalten der relativen Frequenz ermöglicht<sup>26</sup>. Die Granulierung (Abbildung 9, Punkt 6) erlaubt es, den Beobachtungszeitraum zu bestimmen. Im Beispiel ist der Regler auf ‚Tag‘

<sup>23</sup> Wie bereits erwähnt, werden (A) hintereinandergestellte Abfragen in der Grafik zusammengefasst. Ebenso (B) kann durch eine An-/Abwahl einzelner Ergebnisse im Suchverlauf (Abbildung 6, Punkt 2) die Darstellung angepasst werden.

<sup>24</sup> OWIDplusLIVE nutzt das ISO-Datumsformat. Dies besteht aus Jahr (vierstellig) – Monat (zweistellig) – Tag (zweistellig). Daher entspricht 2021-12-18 dem deutschen Datumsformat: 18.12.2021.

<sup>25</sup> Dies ist ersichtlich daran (A) dass die Angabe in ‚(pro Mio. Token)‘ erfolgt und (B) dass der Schieberegler für die ‚Relative Frequenz‘ (siehe Abbildung 9, Punkt 5) aktiviert (blau) ist.

<sup>26</sup> Die relative Frequenz (pro Millionen Token) ist bei Zeitvergleichen angebracht – da hierdurch Schwankungen in der Ergebnismenge pro Tag ausgeglichen werden.



gesetzt, es ist aber auch möglich, Wochen, Monate, Quartale oder Jahre zu wählen – entsprechend größer werden die visualisierten Bereiche. Ein zweiter Regler ‚Glättung‘ (Abbildung 9, Punkt 7) kontrolliert den gleitenden Durchschnitt. Das Einstellungs-Setting Granulierung = ‚Tag‘ und Glättung = 7 bedeutet: Berechne den Frequenzverlauf tageweise und berechne für jeden Tag einen Durchschnitt aus sieben Tagen (im Beispiel: drei Tage davor, der gewählte Tag und drei Tage danach). Falls die Grafik nach mehreren Suchen zu unübersichtlich werden sollte, kann über das Hauptmenü (Abbildung 9, Punkt 8) mit dem Menüpunkt „Neu“ der komplette Such- und Visualisierungsverlauf gelöscht werden. Vorher lohnt aber noch ein Blick auf die Kalender-Visualisierung (Klick auf Abbildung 9, Punkt 9), denn auch wenn die Ausschläge im Zeitverlauf gut sichtbar sind, so lässt sich nur mit etwas Mühe (Bewegungen mit der Maus) die korrekten Datumangaben identifizieren.

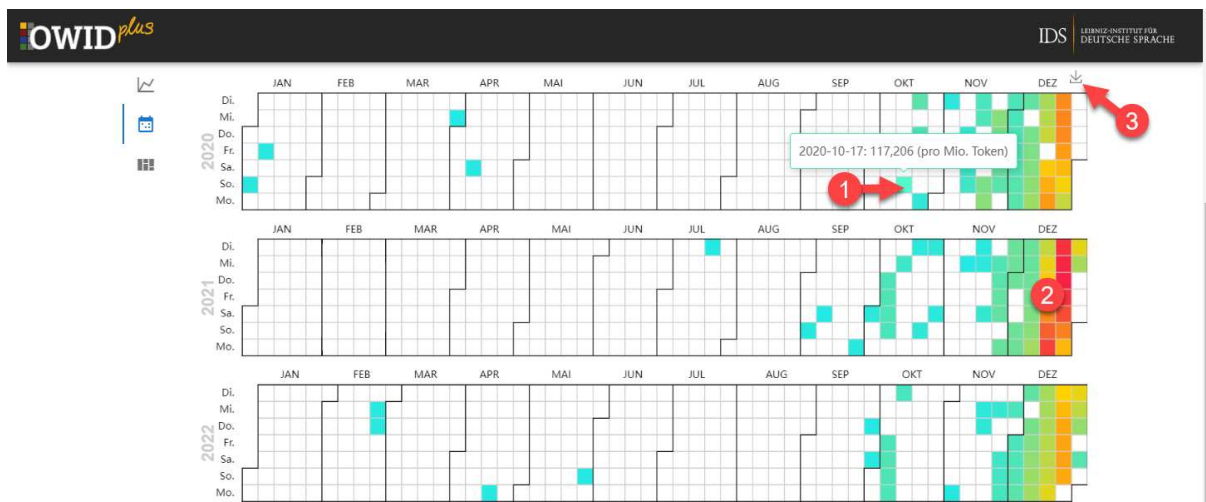


Abbildung 10  
Kalender-Visualisierung zu „ADJA Weihnachten“ mit manuellen Hervorhebungen

Auch die Kalender-Visualisierung ist ähnlich interaktiv – wird der Mauscursor über ein Datumfeld bewegt (Abbildung 10, Punkt 1) werden Informationen angezeigt. Jede Abfrage (auch hintereinandergestellte) wird durch diese Visualisierung summiert. Je ‚wärmer‘ also gelb/orange/rot ein Bereich ist, desto höher ist die Frequenz. Datumfelder ohne einen Fund werden weiß angezeigt. So ist in der Grafik schnell zu erkennen, dass Übereinstimmungen im Korpus nur zu einer bestimmten Zeit vorkommen (jedes Jahr ab ca. Oktober – fast durchgängig im Dezember).

OWIDplusLIVE bietet also eine ganze Reihe von Möglichkeiten nach Bi-Grammen zu suchen. Sowohl für konkrete Suchen (wie z.B. weiße Weihnachten, oranges Auto), als auch für weite Suchen (wie z. B. 1. Lemma = orange, 2. POS = NN). Durch die Nutzung der Visualisierungsfunktion können Ereignisse, die den Frequenzverlauf beeinflussen, entdeckt werden (z. B. suche nach: „zweite Welle“ im Kontext der Corona-Pandemie) und ein Blick in die KorAP-Belege kann eine Analyse substantiieren.

### Fall 3: Gibt es Variationen einer gegebenen Phrase? Welche Bedeutungsunterschiede sind mit der Varianz verbunden?

Zwar kann OWIDplusLIVE aktuell nur für die Recherche von maximal Tri-Grammen (N=3) genutzt werden, jedoch reicht dies meist, um Variationen in Phrasen, Argumentationsmustern und Sprichwörtern zu suchen. Zunächst muss die Suche auf N=3 umgestellt werden (Abbildung 11, Punkt 1).

The screenshot shows the OWIDplus search interface. At the top left is the logo 'OWIDplus' and at the top right is 'IDS LEIBNIZ-INSTITUT FÜR DEUTSCHE SPRACHE'. Below the logo, there is a search bar with the text 'SUCHEN: Einfache Suche'. Underneath, there are three radio buttons for 'Suchfenstergröße: N=1', 'N=2', and 'N=3'. The 'N=3' option is selected and highlighted with a red circle and the number '1'. Below this, there is a section for 'Suchabfrage:' with three input fields. Each field has a 'Suchausdruck (an 1. Position)...' label above it. The first field contains 'POS' and has a dropdown arrow. The second field contains 'KOUS' and has a dropdown arrow. The third field contains 'K\*' and has a dropdown arrow. Below these fields, there are three 'Beispiele (anklicken und ausführen):' links: 'POS = ART / Lemma = unglaublich / POS = NN | Wortform = merkel / POS = KON / POS = NE | POS = APPR / Lemma = menschlich / POS = NN'. At the bottom, there are two buttons: 'ALLE EINGABEN LÖSCHEN' and 'SUCHE AUSFÜHREN'.

Abbildung 11  
Suchmaske für Tri-Gramme mit manuellen Hervorhebungen

Für eine Erprobung wird die Einstellung aller drei Suchslots auf den Layer POS empfohlen (siehe Abbildung 11, Punkt 2) sowie die Werte: KOUS<sup>27</sup>, K\*<sup>28</sup> und ADV<sup>29</sup>. Wird die Suche ausgeführt, so sind die fünf häufigsten Ergebnisse (am 04.05.2023): ‚wenn und aber‘, ‚obwohl oder gerade‘, während als auch‘, ‚ob und inwieweit‘ und ‚nachdem wie lange‘. Für alle diese Ergebnisse können Frequenzen über den Suchverlauf (Abbildung 6) und KorAP-Belege (Abbildung 6, Punkt 8) abgefragt werden.

Das folgende Beispiel soll aber einen Vergleich aller Ergebnisse visualisieren. Dazu ist es notwendig, den Visualisierungsbereich auf die Sankey-Visualisierung umzustellen (Klick auf Abbildung 12, Punkt 1).

<sup>27</sup> KOUS steht im POS-Tagset für eine unterordnende Konjunktion (Bsp.: weil, dass, damit, wenn, ob).

<sup>28</sup> K\* selektiert alle Werte des POS-Tagsets, die mit K beginnen, als KOUI (unterordnende Konjunktion mit ‚zu‘), KOUS (unterordnende Konjunktion), KON (nebenordnende Konjunktion), KOKOM (Vergleichskonjunktion).

<sup>29</sup> ADV steht im POS-Tagset für Adverb.

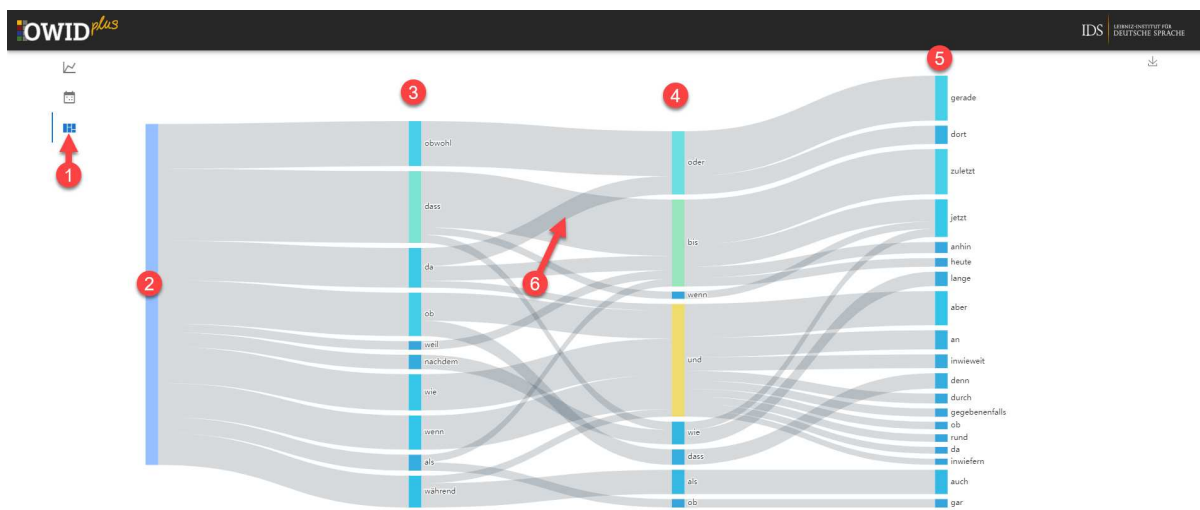


Abbildung 12  
Sankey-Visualisierung für KOUS, K\*, ADV - mit manuellen Hervorhebungen

Die Sankey-Grafik besteht in OWIDplusLIVE immer aus einem Anfangsbereich<sup>30</sup> (Abbildung 12, Punkt 2), gefolgt von allen Slots (da N=3 gewählt, aus drei Positionen). Für jeden Slot (Abbildung 12, Punkt 3, 4 und 5) werden die einzelnen Wortformen aufgeführt. Je größer ein Balken an der jeweiligen Slot-Position dargestellt werden, umso häufiger kommt die Wortform an jener Position vor.

In den Ergebnissen ist *und* an der zweiten Position besonders häufig (siehe Abbildung 12, Punkt 4 – vierter Eintrag von oben / gelb dargestellt). Die grauen Bögen, die die einzelnen Wortformen an den Positionen verbinden, haben unterschiedliche Breiten. Je breiter ein Bogen ist, desto häufiger ist die jeweilige Abfolge. Wird der Mauszeiger über einen Bogen bewegt, werden Informationen zur Häufigkeit angezeigt. Die Bögen ermöglichen eine grobe optische Abschätzung. An der ersten Position ist: *dass* der häufigste Eintrag (Abbildung 12, Punkt 3 – zweiter Eintrag von oben). Von diesem gehen drei (nach rechts) Bögen ab. Der breiteste Bogen und damit die häufigste Folge ist: *bis* (Abbildung 12, Punkt 4 – zweiter Eintrag von oben). Auf *bis* folgt am häufigsten *zuletzt*. Die komplette Abfolge ‚*dass bis zuletzt*‘ kommt auch in der Ergebnismenge vor. Dies muss für Tri-Gramme (N=3) nicht zwingend der Fall sein, denn die Sankey-Visualisierung zeigt nur die häufigsten Positionen und deren Verbindungen zwischen zwei Positionen.

Werden via KorAP Belege zu ‚*dass bis zuletzt*‘ gesucht, so ergibt sich z.B. folgende zufällige Ausgabe:

„Vieles ist geklärt, aber nicht alles - und das liegt nicht nur daran	<b>dass bis zuletzt</b> BOZ16/APR/00762	alles noch unter Finanzierungsvorbehalt steht.
„Ihr Switchen zwischen Realität und Phantasiewelt ist so subtil	<b>dass bis zuletzt</b> BUN06/MAR/00995	eine Frage offen bleibt: Wie viel Madonna erträgt die Liebe?
„Eine Aufführung, der man anmerkt	<b>dass bis zuletzt</b> BUN04/MAR/00419	improvisiert und experimentiert worden ist, die aber ebenso wie letztes Jahr «Groundings» ze...
...ndere durch schwache US-Wirtschaftsdaten dominierte, trübe Umfeld führt in jedem Fall dazu	<b>dass bis zuletzt</b> BOZ11/AUG/00422	sehr positiv gestimmte Anlagestrategen nun ihre Erwartungen für das zweite Halbjahr korrigie...
	<b>Dass bis zuletzt</b> BZE15/NOV/01199	Bundesliga-Funktionäre betonten, wie zuverlässig Niersbach war, spricht menschlich für ihn.

Abbildung 13  
Fünf ausgewählte KorAP-Belege zu ‚*dass bis zuletzt*‘

<sup>30</sup> Dieser ist nur funktional und hat für die Interpretation keine Bedeutung.

### 3. Fazit

Wie viele Korpora und korpusgestützte Ressourcen, so sind auch die hier gezeigten Ressourcen sehr stark an der Korpuslinguistik und damit primär an der Sprachwissenschaft orientiert. Der Artikel versuchte zu zeigen, welche Ressourcen online verfügbar und leicht nutzbar sind. Das DWDS eignet sich als erste Anlaufstelle, die auch DaF/Z-Lerner\*innen mit etwas Vorkenntnis ab B1-Niveau nutzen können (ggf. begleitet durch die Lehrkraft). OWID und OWIDplus bietet eine große Bandbreite an sprachwissenschaftlichen Ressourcen, die sich auch im Deutschunterricht einsetzen lassen, bzw. einer Lehrkraft interessantes sprachwissenschaftliches Hintergrundwissen bieten (z.B. Diskurswörterbücher, Sprichwörterbücher, etc.). OWIDplusLIVE ist die erste, vollständig frei zugängliche Ressource für tagesaktuelle Analysen deutscher Pressesprache und direkt verknüpft mit KorAP (hier ist eine kostenlose Registrierung erforderlich) zur Volltextrecherche. Gerade wenn tagesaktuelle Ereignisse Unterrichtsgegenstand sein sollen, bietet sich eine Nutzung an<sup>31</sup>.

Auch wenn die Nutzung durch webbasierte Ressourcen den Unterrichtseinsatz vereinfachen<sup>32</sup>, indem diese ohne Installation mit jedem Computer<sup>33</sup> nutzbar sind, so ist doch oft eine Einarbeitung nötig (siehe Literaturempfehlungen oben) und es erfordert einiges an Kreativität, um Beispiele sowie daraus resultierende Materialien zu generieren. Gerade fortgeschrittene DaF/Z-Lerner\*innen ab C1-Niveau können von korpusbasierten Ressourcen profitieren, wenn sie lernen, diese für eigene Recherchen zu nutzen. Dazu ist es notwendig, dass die Lehrkraft die notwendigen Grundkenntnisse, z.B. zu den eingesetzten Ressourcen vermittelt und ggf. notwendige technische Voraussetzungen (WLAN, Endgeräte, etc.) schafft. Die Lerner\*innen können dann einzeln oder in einer Kleingruppe konkreten Fragestellungen nachgehen (je nach Unterrichtsform<sup>34</sup>). Hierbei ist darauf zu achten, dass es einer gewissen Einarbeitungszeit bedarf (Computer starten, sich mit den Ressourcen vertraut machen, ggf. gute Beispiele für die Fragestellung zu suchen). Ideal ist es daher, mehrere Ressourcen innerhalb einer längeren Zeitspanne, z.B. Projektwoche, vorzustellen und die Lerner\*innen selbstständig damit arbeiten zu lassen. Für den Unterrichtseinsatz ist es zudem oft sinnvoll, die Planung nicht von der eigentlichen Ressource her zu denken (auch wenn diese vorgestellt und eingeführt werden muss), sondern vielmehr das Ergebnis, bzw. die Ergebnispräsentation in den Blick zu nehmen. Einfach nur ein Online-Wörterbuch für wenige Minuten oder ein paar Unterrichtsstunden abzufragen, schafft nur ein geringes Bewusstsein dafür, wie man diese Ressource zum eigenen Lernfortschritt verwenden kann. Vom Ziel gedacht, können Lerner\*innen nicht nur eine Gruppendiskussion oder ein Kurzreferat führen, sondern z.B. auch ein Plakat gestalten, einen Podcast aufnehmen oder sogar ein (YouTube, TikTok, etc.) Video produzieren. Dies motiviert nicht nur, sondern macht es auch erforderlich, dass sich die Lerner\*innen intensiver mit Frage-/Problemstellung, der Ressource und der Analyse auseinandersetzen müssen. Da zur Problemlösung vielfältige Kompetenzen (z. B. Lesen, Sprechen, Hören ebenso wie der Umgang mit digitalen Medien) nötig sind, entstehen in der Praxis oft produktive Diskussionen in den Kleingruppen oder auch im Plenum.

---

<sup>31</sup> Für einen Blick auf lang zurückliegende Ereignisse, sei hier auf die Wortverlaufskurven des DWDS verwiesen. <https://www.dwds.de/d/resources#wortverlauf> (11.11.2023).

<sup>32</sup> Viele Korpus-Tools erfordern erweitertes Hintergrundwissen und ggf. entsprechende Vorbereitungen der Lehrkraft. Siehe hierzu auch die Praxisberichte z.B. von Bubenhofer (2011) und Dipper (2011).

<sup>33</sup> Unabhängig vom Betriebssystem z.B. MacOS, Linux oder Windows.

<sup>34</sup> Z.B.: Feste Themenvergabe durch die Lehrperson; Rahmenthema wie z.B. Argumente oder grammatische Fragen – die Gruppen entwickelt davon abgeleitete eigene Fragestellungen ggf. in Diskussion mit Lehrperson oder Plenum; bei besonders fortgeschrittenen Lerner\*innen ab C1-Niveau kann auch nur die Ressource vorgestellt werden und die Gruppen sollen damit experimentieren und eine eigene Fragestellung erarbeiten.

## Literatur und Ressourcen

- Barbaresi, Adrien (2022): Webmonitor. <https://www.dwds.de/d/korpora/webmonitor> (11.11.2023).
- Bubenhof, Noah (2009): *Sprachgebrauchsmuster: Korpuslinguistik als Methode der Diskurs- und Kultur-analyse*. Sprache und Wissen. Bd. 4. Berlin: De Gruyter.
- Bubenhof, Noah (2011): Korpuslinguistik in der linguistischen Lehre: Erfolge und Misserfolge. In: *Journal for Language Technology and Computational Linguistics* 26: 1, 141-156. doi:10.21248/jlcl.26.2011.141.
- Davies, Mark (2013): Corpus of News on the Web (NOW). <https://corpus.byu.edu/now/> (11.11.2023).
- Dipper, Stefanie (2011): Digitale Korpora in der Lehre - Anwendungsbeispiele aus der Theoretischen Linguistik und der Computerlinguistik. In: *Journal for Language Technology and Computational Linguistics* 26: 1, 81-95.
- Dudenredaktion (2009): *Duden – Die Grammatik: Unentbehrlich für richtiges Deutsch*. 8., überarb. Aufl. Mannheim et al: Dudenverlag.
- DWDS (2023): DWDS – Das Digitale Wörterbuch der deutschen Sprache. <http://dwds.de/> (11.11.2023).
- Flinz Carolina (2021): Korpora in DaF und DaZ: Theorie und Praxis. In: Flinz Carolina / Britta Hufeisen: *Korpora in DaF und DaZ: Theorie und Praxis*. (Themenheft). Zeitschrift für Interkulturellen Fremdsprachenunterricht. Didaktik und Methodik im Bereich Deutsch als Fremdsprache. Jahrgang 26, N.1 (April 2021), 1-43. <https://zif.tu-journals.ulb.tu-darmstadt.de/article/id/3321/> (18.11.2023).
- Heyer, Gerhard / Quasthoff, Uwe / Wittig, Thomas (2006): *Text Mining: Wissensrohstoff Text - Konzepte, Algorithmen, Ergebnisse*. Herdecke, Bochum: W3L-Verlag.
- Kupietz, Marc (2020): *Deutsches Referenzkorpus – DeReKo-2020-I*. Mannheim: Leibniz-Institut für Deutsche Sprache. <http://hdl.handle.net/10932/00-04B6-B898-AD1A-8101-4> (11.11.2023).
- Lemnitzer, Lothar / Zinsmeister, Heike (2015): *Korpuslinguistik: eine Einführung*. 3., durchgesehene und aktualisierte Auflage. Tübingen: Narr Verlag.
- OWID (2023): OWID – Online-Wortschatz-Informationssystem Deutsch. <http://www.owid.de/> (11.11.2023).
- Perkuhn, Rainer / Keibel, Holger / Kupietz, Marc (2012): *Korpuslinguistik*. UTB Sprachwissenschaft 3433. Paderborn: Fink.
- Schiller, Anne / Teufel, Simone / Stöckert, Christine / Thielen, Christine (1999): *Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset)*. Universität Stuttgart, Universität Tübingen. <http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf> (11.11.2023).
- Vogel, Friedemann / Deus, Fabian / Knobloch, Clemens / Rüdiger, Jan Oliver / Schmallenbach, Joline / Schölzel, Hagen / Tripps, Felix / Bäumer, Benjamin / Weber, Susanna / Wilton, Antje (2021): [www.Diskursmonitor.de](http://www.Diskursmonitor.de) – gemeinschaftlich erarbeitete Online-Plattform zur Aufklärung und Dokumentation strategischer Kommunikation.
- Wolfer, Sascha / Koplenig, Alexander / Michaelis, Frank / Müller-Spitzer, Carolin (2020): Tracking and analyzing recent developments in German-language online press in the face of the coronavirus crisis: cOWIDplus Analysis and cOWIDplus Viewer. In: *International Journal of Corpus Linguistics* 25: 347-359. <https://doi.org/10.1075/ijcl.20078.wol>.
- Wolfer, Sascha (2020): cOWIDplus. <https://www.owid.de/plus/cowidplus2020/> (11.11.2023).



**Biographische Notiz:** Jan Oliver Rüdiger ist wissenschaftlicher Mitarbeiter am Leibniz-Institut für Deutsche Sprache in Mannheim. Der Schwerpunkt seiner Tätigkeit ist die Konzeption und Umsetzung von hochspezialisierten Softwarelösungen für die germanistische Sprachwissenschaft und die Digital Humanities.

**Kontaktanschrift:**

Jan Oliver Rüdiger  
Leibniz-Institut für Deutsche Sprache  
R 5, 6-13  
68161 Mannheim  
Deutschland  
[ruediger@ids-mannheim.de](mailto:ruediger@ids-mannheim.de)

