

Möglichkeiten und Grenzen korpuslinguistischer Analysen von Synonymen am Beispiel der Derivate auf *-ierung* und *-ion*¹

Martin MOSTÝN

University of Ostrava
martin.mostyn@osu.cz

ABSTRACT

Possibilities and boundaries of corpus linguistic synonym analysis – the example of derivatives with *-ierung* and *-ion*

The focus of this study is on deverbal derivatives of verbs in *-ieren*, which can occur both with the suffix *-ung* and with *-ion*. The competing variants such as *Assimilierung/Assimilation*, *Modifizierung/Modifikation*, *Stimulierung/Stimulation* are used partly synonymously, partly differentiated in their meaning. Using selected examples, various possibilities and limitations of a corpus-based analysis of these parallel formations will be discussed while results obtained from different language corpora will be compared. In particular, the German reference corpus (with the analysis tool Cosmas II and the co-occurrence analysis with the help of the co-occurrence database CCDB from Cyril Belica), the reference and newspaper corpora of the Digital Dictionary of the German Language (DWDS) (with the analysis tool for co-occurrence profiles), the vocabulary portal (Deutscher Wortschatz) of the University of Leipzig with co-occurrence profiles and, for comparison, the Google Books corpus (with the analysis tool Google Books NGram Viewer) will be used. In doing so, the question is pursued as to whether the results obtained agree with regard to the occurrence and co-occurrence of both competing formations on *-ierung* and *-ion*, which were determined using different corpora, and to what extent these corpora are suitable for the analysis of such near synonyms.

KEYWORDS

synonyms, deverbal derivatives, synchronic foreign word formation, co-occurrence, corpus linguistics

1 Die hier präsentierten Überlegungen stützen sich auf die im Rahmen meines Habilitationssprojekts (‘Deverbale Derivate von Verben auf *-ieren* – eine metalexikografische und korpuslinguistische Analyse’ (s. Mostýn 2022b) durchgeführten Analysen.

1. Einleitend zur Problematik

In der deutschen Gegenwartssprache bestehen zahlreiche Parallelbildungen mit den Suffixen *-ung* und *-ion* nebeneinander, wie z. B. *Diversifizierung/Diversifikation*, *Evaluierung/Evaluation*, *Kodifizierung/Kodifikation*, *Ionisierung/Ionisation*, *Nostrifizierung/Nostrifikation*, *Ratifizierung/Ratifikation*, *Stimulierung/Stimulation* u. v. a., die sich ausgehend von der synchronen Fremdwortbildungsanalyse² auf die gleiche Derivationsbasis – die Verben auf *-ieren* – zurückführen lassen und dementsprechend als deverbale Derivate von diesen Verben betrachtet werden können. Die oben genannten Beispiele von Derivaten stellen sog. Wortbildungskonkurrenzen dar, da sie formale und funktionale Gemeinsamkeiten aufweisen und in den meisten Kontexten ohne Bedeutungsunterschied einander substituieren können (zum Begriff „Wortbildungskonkurrenz“ s. Stricker 2002). Hinsichtlich paradigmatischer Relationen der Synonymie werden solche gleichbedeutenden Ausdrücke als „absolute Synonyme“ bezeichnet (genannt auch „echte Synonyme“, „Signifié-Synonyme“, „Vollsynonyme“, „totale Synonyme“, „full synonyms“) (vgl. Hümmer 2007:28; Hahn 2002:36 ff.). Dies spiegelt sich ebenfalls bei der lexikografischen Bearbeitung entsprechender Lemmata in einsprachigen Wörterbüchern wider, indem im Interpretament in der Bedeutungsparaphrase häufig lediglich eine Infinitivkonversion, die zugleich eine Konkurrenzform darstellt, angeführt wird, wie es beispielsweise bei *Evaluierung/Evaluation* – ‚das Evaluieren‘ oder *Stimulierung/Stimulation* – ‚das Stimulieren‘ der Fall ist (vgl. ‚Duden Online-Wörterbuch‘, im Folgenden ‚DOW‘). Bisweilen werden unter synonymen Ausdrücken von *-ung*-Bildungen auch entsprechende Parallelbildungen auf *-ion* angegeben, und umgekehrt, wie z. B. an den konkurrierenden Derivaten *Ratifikation* (Syn.: *Bestätigung*, *Inkraftsetzung*, *Ratifizierung*) und *Ratifizierung* (Syn.: *Bestätigung*, *Inkraftsetzung*, *Ratifikation*) zu sehen ist (s. ‚Digitales Wörterbuch der deutschen Sprache‘, im Folgenden ‚DWDS‘), oder als Bedeutungsparaphrase wird lediglich die Konkurrenzform herangezogen: *Ratifizierung* = *Ratifikation* (s. ‚Brockhaus Wahrig Deutsches Wörterbuch 2011:1199). Aus syntagmatischen Relationen dieser und anderer ähnlicher Derivate, die sich in ihren Konkurrenzprofilen manifestieren, geht dennoch hervor, dass die Bildungen auf *-ierung* und *-ion* in verschiedenen Kontexten häufig durch eine unterschiedliche Distribution und zugleich durch eine unterschiedliche funktionale Auslastung gekennzeichnet sind, wobei bisweilen auch areale Unterschiede im Hinblick auf ihren Gebrauch in den DACH-Ländern festzustellen sind (s. dazu Mostýn 2017, 2019a, 2019b, 2020, 2022a, ausführlich in 2022b).

2 Zur synchronen Fremdwortbildungsanalyse s. u. a. Müller (2005:22 ff.; 2015:1620 ff.), Seifert (2005:227 ff.).

Manche Parallelbildungen auf *-ierung* und *-ion* zeichnen sich hingegen durch eine partielle Substituierbarkeit aus, die zum Teil auf ihren unterschiedlichen transformationellen Wert zurückzuführen ist (zum transformationellen Wert deverbalen Substantive s. Wellmann 1975). So sind beispielsweise die Derivate wie *Dokumentierung/Dokumentation*, *Injizierung/Injektion*, *Komprimierung/Kompression*, *Legitimierung/Legitimation*, *Modifizierung/Modifikation*, *Rekonstruierung/Rekonstruktion*, *Tradierung/Tradition* als Nomina actionis in der deutschen Sprache als Synonyme belegt, als Nomina acti bzw. als Zustandsabstrakta sind dagegen nur entsprechende Bildungen auf *-ion* gebräuchlich. Auch bei Derivaten, die den gleichen transformationellen Wert aufweisen und beispielsweise als Nomina actionis Verwendung finden, treten mitunter erhebliche Unterschiede im Hinblick auf ihren Gebrauch in verschiedenen Lesarten sowie in Bezug auf ihre Verträglichkeit mit verschiedenen Kookkurrenzpartnern zum Vorschein, was sich etwa bei Derivaten wie *Absolvierung/Absolution*, *Degradierung/Degradation*, *Immobilisierung/Immobilisation*, *Kassierung/Kassation*, *Konzertierung/Konzertation*, *Raffinierung/Raffination*, *Reduzierung/Reduktion* bemerkbar macht (s. dazu ausführlich in Mostýn 2022b). Unter paradigmatischem Gesichtspunkt werden solche Ausdrücke „partielle Synonyme“ genannt (auch „Sem-Synonyme“, „Quasi-Synonyme“, „Pseudo-Synonyme“, „sense synonyms“) (vgl. Hümmer 2007:28; Hahn 2002:36 ff.).

Das Zusammenspiel von Ähnlichkeiten und Unterschieden bei der Verwendung bedeutungsgleicher Ausdrücke in verschiedenen Kontexten wird im Zusammenhang mit den sog. „nahen Synonymen“ diskutiert (engl. „near synonyms“) (s. dazu Marková 2012:26; Storjohann 2006:11). Nahe Synonyme (auch „Plesionyme“ genannt) stellen Ausdrücke dar, die „über eine bestimmte Menge gemeinsamer salienter Merkmale verfügen“ (Storjohann ebd.), aber bei denen im Hinblick auf ihren Gebrauch in verschiedenen Kontexten deutliche Kontraste hervortreten. Im Zusammenhang mit den hier behandelten Parallelbildungen heißt das, dass die Verwendung der Bildung auf *-ierung* bzw. auf *-ion* in Kontexten, in denen diese nicht austauschbar sind bzw. in denen entweder die erstere oder die letztere dominiert, häufiger zu verzeichnen ist als das Vorkommen beider Derivate in gemeinsamen Kontexten. Storjohann deutet in diesem Zusammenhang darauf hin, dass die Grenze zwischen semantischer Nähe und semantischer Distanz korpusbasiert überprüft werden kann (ebd.). Dazu sind insbesondere Kookkurrenzanalysen geeignet, die mithilfe verschiedener Korpustools durchgeführt werden können.

Bevor detaillierte Kookkurrenzanalysen deverbalen Derivate von Verben auf *-ieren*, die Aufschluss über ihre Distribution in verschiedenen Kontexten bzw. Lesarten und über den Grad an Synonymie geben, durchgeführt werden können, ist es notwendig, zunächst ihre Okkurrenz zu untersuchen. Dabei gilt

es herauszufinden, bei welchen Verben auf *-ieren* Derivate a) weder auf *-ion* noch auf *-ierung*, b) nur auf *-ion*, c) nur auf *-ierung*, und schließlich sowohl auf *-ion* als auch auf *-ierung* in Korpora belegt sind. Im Folgenden werden an ausgewählten Beispielen verschiedene Möglichkeiten und Grenzen einer korpusbasierten Analyse der Okkurrenz und Kookkurrenz dieser Parallelbildungen erörtert, wobei auch auf einige Problemfälle bei Korpusrecherchen hingewiesen wird. Dabei werden Ergebnisse verglichen, die anhand verschiedener Sprachkorpora ermittelt werden. Herangezogen werden insbesondere das Deutsche Referenzkorpus (mit dem Analysetool Cosmas II und der Kookkurrenzanalyse mit Hilfe der Kookkurrenzdatenbank CCDB, die Referenz- und Zeitungskorpora des Digitalen Wörterbuchs der deutschen Sprache (mit dem Analysetool für Kookkurrenzprofile – DWDS-Wortprofile), das Wortschatz-Portal der Universität Leipzig mit Kookkurrenzprofilen und zum Vergleich auch das Google Books-Korpus (mit dem Analysetool Google Books NGram Viewer). Dabei wird der Frage nachgegangen, ob die gewonnenen Ergebnisse in Bezug auf die Okkurrenz und Kookkurrenz beider konkurrierenden Bildungen auf *-ierung* und *-ion*, die anhand verschiedener Korpora ermittelt wurden, übereinstimmen und inwiefern diese Korpora zur Analyse solcher Synonyme geeignet sind, um schließlich ein optimales methodisches Vorgehen zu entwerfen, das die Relation zwischen ihrer Form und Funktion aufzudecken hilft.

2. Analyse der Okkurrenz deverbaler Derivate von Verben auf *-ieren*

2.1. Identifizierung der Formen

Im ersten Schritt gilt es, einzelne Bildungen auf *-ion* und *-ierung* zu identifizieren. Möchte man sich einen ersten Überblick über die Anzahl der Substantive auf *-ierung* und *-ion* verschaffen, kann man u. a. das Deutsche Referenzkorpus (im Folgenden DeReKo) des IDS Mannheim heranziehen. Die geschriebenen Korpora des DeReKo umfassen verschiedene Archive mit insgesamt 53 Mrd. laufenden Wortformen (Stand 03/2023, URL 1). Die umfangreichsten Archive stellen die vier Archive der geschriebenen Sprache dar, die mit den Archivsignen W, W2, W3 und W4 versehen sind. Die Archive sind im Hinblick auf ihren Umfang vergleichbar, wie aus der folgenden Tabelle³ ersichtlich ist:

3 In der Tabelle ist jeweils die Korpusgröße aller öffentlichen Korpora mit Neuakquisitionen angegeben.

DeReKo-W öffentlich	10.764.331.064
DeReKo-W2 öffentlich	12.310.699.925
DeReKo-W3 öffentlich	12.077.307.005
DeReKo-W4 öffentlich	11.605.179.020
Summe der Tokens	46.757.517.014

TAB. 1: Umfang der vier Archive der geschriebenen Sprache des DeReKo (Stand 03/2023)

Sie unterscheiden sich jedoch bezüglich ihrer Korpuszusammensetzung. Am stärksten repräsentiert sind in den vier Archiven eindeutig verschiedene Presse-texte, die einen beträchtlichen Teil des öffentlichen Sprachgebrauchs abbilden. Das Archiv W (im Folgenden DeReKo-W) umfasst neben Zeitungstexten aus auflagenstarken, großen Zeitungen, die den größten Anteil ausmachen, u. a. auch ausgewählte Zeitschriftentexte aus diversen thematischen Bereichen, ferner belletristische Texte, einige Fachtexte, aber auch Sachbücher und Ratgeber, Biografien, Agenturmeldungen, Protokolle aus Plenarsitzungen verschiedener Landtage und Wikipediaartikel aus dem Jahre 2011 mit einschlägigen Diskussionen. In den Archiven W2, W3 und W4 sind überwiegend Zeitungstexte aus kleineren regionalen Zeitungen zu finden, darüber hinaus sind im W2-Archiv noch einige Lifestyle- und Fachzeitschriften enthalten. Der obigen Tabelle ist zu entnehmen, dass das Archiv W2-öffentlich zurzeit am umfangreichsten ist. Möchte man verlässliche Daten für eingehende Kookkurrenzanalysen gewinnen, sind sehr große Korpora erforderlich, die eine möglichst hohe Anzahl an Treffern für verschiedene Kookkurrenzen und verschiedene Kontexte liefern. Aus diesem Grund spielt die Korpusgröße bei der Korpusauswahl eine sehr wichtige Rolle. Dennoch sind bei der Wahl des zu analysierenden Materials auch andere Korpuseigenschaften in Betracht zu ziehen, wie am Beispiel der Vorkommenshäufigkeit der Parallelbildungen *Stimulierung/Stimulation* und *Perzipierung/Perzeption* in den vier Archiven des DeReKo demonstriert wird. In der folgenden Tabelle sind die absolute und die relative Frequenz (pro Million Wörter) (im Folgenden $f_{(a)}$ und $f_{(r)}$) eingetragen:

	<i>Stimulierung</i>		<i>Stimulation</i>		<i>Perzipierung</i>		<i>Perzeption</i>	
	f _(a)	f _(r)	f _(a)	f _(r)	f _(a)	f _(r)	f _(a)	f _(r)
DeReKo-W	3.299	0,306	5.977	0,555	2	0.0002	748	0,0695
DeReKo-W2	2.509	0,204	5.313	0,4316	0	0	223	0,0181
DeReKo-W3	979	0,0782	3.527	0,2920	0	0	110	0,0091
DeReKo-W4	1.321	0,114	3.388	0,292	0	0	158	0,0136

TAB. 2: Die f_(a) und f_(r) der Parallelbildungen *Stimulierung/Stimulation* und *Perzipierung/Perzeption* in den Archiven W1, W2, W3 und W4 des DeReKo (Stand 03/2023)

Aus den Frequenzwerten beider Derivate auf *-ierung* sowie auf *-ion* ergibt sich, dass ihre Vorkommenshäufigkeit im Archiv W am höchsten ist, auch wenn dieses Archiv unter den vier Archiven am kleinsten ist. Dies ist auf die unterschiedliche Zusammensetzung der einzelnen Archive im Hinblick auf enthaltene Textsorten und behandelte Themen zurückzuführen, denn im DeReKo-W sind vornehmlich wegen der Einbeziehung der Wikipedia-Artikel deutlich mehr Fachtexte vertreten als in den anderen drei Archiven. Da die Bildungen auf *-ierung* und *-ion* sehr häufig im fachsprachlichen Kontext auftreten, sind sie dementsprechend im DeReKo-W auch häufiger zu verzeichnen als in den Archiven W2, W3 und W4, bei denen der Anteil an Fachtexten geringer ist. Des Weiteren ist an den Frequenzen ersichtlich, dass während bei *Stimulierung/Stimulation* im öffentlichen Sprachgebrauch beide Parallelbildungen gebräuchlich sind, ist *Perzipierung* gegenüber *Perzeption* nur äußerst selten belegt und nur im DeReKo-W zu verzeichnen. Da im Rahmen der Analyse der Okkurrenz deverbaler Derivate von Verben auf *-ieren* herauszufinden ist, welche Formen im deutschen Sprachgebrauch überhaupt belegt sind (auch wenn sie selten sind), ist unter den vier Archiven das DeReKo-W auszuwählen, denn je höher die Trefferzahl ist, desto verlässlicher und aussagekräftiger sind die Ergebnisse von Korpusrecherchen (vgl. dazu Mostýn 2019b).

Wie bereits erwähnt, sind sehr große Korpora wie das DeReKo-W für Okkurrenzanalysen bestens geeignet, da sie im Rahmen verschiedener Suchanfragen eine möglichst hohe Trefferzahl bieten. Dennoch kann sich die Korpusgröße bei Korpusrecherchen mitunter auch als ein Hindernis erweisen: Möchte man beispielsweise im Rahmen der Analyse der Okkurrenz korpusbasiert vorgehen und die absolute sowie relative Frequenz der Derivate auf *-ion* in diesem Korpus ermitteln, kann man sich zunächst einer einfachen Suchanfrage – **ion* – bedienen. Bei dieser Suchanfrage muss jedoch in Kauf genommen werden, dass sie nicht nur deverbale Derivate auf *-ion*, sondern auch denominalen Derivate wie *Desillusion*, *Exkursion*, *Institution*, *Translation*, deadjektivische Derivate wie *Dis-*

ktion, Devotion, ferner Simplizia, denen im Deutschen kein Verb auf *-ieren* zugrunde liegt, wie *Emotion, Illusion, Mission, Station, Satisfaktion* und jede Menge von Determinativkomposita liefert, deren Determinatum ein *-ion*-Substantiv darstellt.⁴ Insgesamt ergibt die oben erwähnte Suchanfrage 740 820 verschiedene Wortformen (Stand 03/2023). Aufgrund ihrer sehr hohen Anzahl wird die Menge der einsehbaren Formen automatisch auf 200 000 reduziert. Auch die absolute bzw. relative Frequenz dieser Substantive ist im DeReKo-W aktuell nicht zu ermitteln, denn aufgrund der hohen Anzahl an Wortformen wird im Cosmas II jedes Mal eine Fehlermeldung angezeigt.

Das Korpustool Cosmas II bietet neben Suchanfragen, die eine Suche nach Ausdrücken mit einer bestimmten Lautkombination (nicht nur) im Auslaut ermöglichen, auch eine Suche nach Wortformen mit ausgewählten Präfixen und Suffixen. Unter den Fremdsuffixen sind z. B. *-ismus, -ist, -ität, -itis*, zu nennen, die entsprechende Suchanfragen ermöglichen (z. B. *♣-ist* – Option – Lemmatisierung – Sonstige Wortbildungsformen). Suchanfragen mit dem Suffix *-ion* sind jedoch zurzeit nicht möglich. Die Frequenzwerte der *-ion*-Substantive können lediglich in einem kleineren Korpus ermittelt werden, beispielsweise beträgt die $f_{(a)}$ aller *-ion*-Substantive im Singular im Wikipedia-Korpus wpd11 (Stand 10/2011) 2 975 440 Tokens bei 141 592 verschiedenen Wortformen, ihre $f_{(r)}$ beläuft sich auf 5 176,8 pMW, während im Korpus Berliner Morgenpost (1997–2022) bei 56 146 verschiedenen Wortformen insgesamt 1 457 821 Substantive auf *-ion* zu verzeichnen sind (Stand 03/2023). Dabei beträgt die $f_{(r)}$ 3 502,7 pMW. Die deutlich niedrigere $f_{(r)}$ im letztgenannten Korpus ist auf einen niedrigeren Fachlichkeitsgrad dieses Korpus zurückzuführen.

Aufgrund des gemeinsamen Vorkommens des exogenen Verbalisierers *-ier-* (mit seinen verwandten Formen *-isier-*, *-ifizier-* und *-izier-*) und des indigenen Suffixes *-ung* lassen sich die Bildungen auf *-ierung* im Gegensatz zu den Bildungen auf *-ion* eindeutig deverbalen Derivaten zuordnen und erfordern dementsprechend keine spezielle morphosyntaktische Annotation. Die Suchanfrage **ierung* liefert im DeReKo-W insgesamt 126 693 Substantive auf *-ierung* in der Singularform. In diesem Fall ist die Anzahl an Wortformen nicht so hoch, wie bei den *-ion*-Substantiven, so dass die $f_{(a)}$ sowie die $f_{(r)}$ der Derivate erfasst werden können. Die erstere beträgt im DeReKo-W 11 624 254 Tokens, was der $f_{(r)}$

4 Eine Differenzierung entsprechender *-ion*-Substantive nach Wortbildungskriterien (wie z. B. I. Simplex – II. Fremdwortbildungsprodukt: IIa) Derivat, IIb) Determinativkompositum) und nach ihrer Derivationsbasis (IIa1) deverbales/IIa2) denominales/IIa3) deadjektivisches Derivat) würde eine fortgeschrittene morphosyntaktische Annotation erforderlich machen. Diese ist zurzeit aufgrund einer sehr hohen Komplexität solcher Suchanfragen schwierig zu bewerkstelligen und bleibt weiterhin ein Desideratum der künftigen Korpuslinguistik.

von etwa 1 079,9 pMW entspricht (Stand 03/2023). An den Frequenzen macht sich bemerkbar, dass die *-ion*-Substantive im DeReKo-W deutlich häufiger zu verzeichnen sind als die Bildungen auf *-ierung*.

Wie hier bereits an verschiedenen Beispielen demonstriert wurde, hängt die Okkurrenz der Derivate auf *-ierung* und *-ion* und damit zusammenhängende Frequenzwerte nicht nur von der Korpusgröße, sondern auch von der Korpuszusammensetzung ab. Aus diesem Grund ist es geeignet, einschlägige Analysen anhand von mehreren Korpora durchzuführen, und anschließend die Ergebnisse einem Vergleich zu unterziehen, ob diese statistisch signifikant sind (s. u.). Was die Ermittlung der Anzahl der Substantive auf *-ion* und *-ierung* anbelangt, können als Vergleichskorpora z. B. die Referenz- und Zeitungskorpora des ‚DWDS‘ herangezogen werden (im Folgenden DWDS-RZ). Die DWDS-Korpora umfassen in ihrer Gesamtheit etwa 49 Mrd. Tokens, der Korpusumfang der DWDS-RZ beläuft sich zurzeit auf mehr als 1,3 Mrd. laufende Wortformen (Stand 03/2023, vgl. ‚DWDS‘; s. dazu auch Didakowski/Geyken 2012; Geyken 2011 und die Publikationsliste unter ‚DWDS‘ s. URL 2). Die Referenz- und Zeitungskorpora umfassen neben Zeitungstexten auch Blogs, Webkorpora, einige Fachzeitschriften und darüber hinaus auch verschiedene Spezialkorpora wie das DDR-Korpus, das Berliner Wende-Korpus, politische Reden und auch historische Texte. Mit der gleichen Suchanfrage wie im DeReKo (**ion*) kann die Gesamtzahl der Substantive auf *-ion* im Singular in den DWDS-RZ ermittelt werden, allerdings ohne eine Wortformenliste, was im Vergleich mit dem Korpusinstrument Cosmas II eine wesentliche Einschränkung im Bereich der Suchfunktionen darstellt. Die $f_{(a)}$ der Substantive auf *-ion* beläuft sich auf 4 093 943 Tokens, was einer $f_{(r)}$ von rund 3 071,2 pMW entspricht (Stand 03/2023). Im Vergleich mit den oben erwähnten Korpora des DeReKo-W weisen die *-ion*-Substantive in den DWDS-RZ die niedrigste Frequenz auf, auch wenn in diesem Korpus verschiedene Fachtexte vertreten sind. Hier manifestieren sich möglicherweise auch die Auswirkungen der Textauswahl. Das DWDS verfügt über verschiedene Korpustools, die eine Visualisierung von Ergebnissen in Form von Verlaufskurven, Clouds und auch verschiedene Kookkurrenzanalysen (s. u.) ermöglichen. So lässt sich beispielsweise die Okkurrenz der *-ion*-Substantive in den DWDS-RZ in verschiedenen Textklassen visualisieren – eine Funktion, die wiederum in Cosmas II fehlt und deren Implementation im Rahmen von Korpusrecherchen im DeReKo begrüßenswert wäre. Vergleicht man die Verlaufskurven der *-ion*-Substantive in verschiedenen Textklassen der DWDS-RZ in der folgenden Abbildung, ist das höchste Vorkommen erwartungsgemäß in der Wissenschaftssprache zu verzeichnen, die sich durch eine hohe terminologische Dichte auszeichnet, was mit einem häufigeren Gebrauch verschiedener Termini auf *-ion* einhergeht, und am nied-

rigsten in der Belletristik, für die eher der Verbalstil und ein niedriger Abstraktionsgrad kennzeichnend sind:

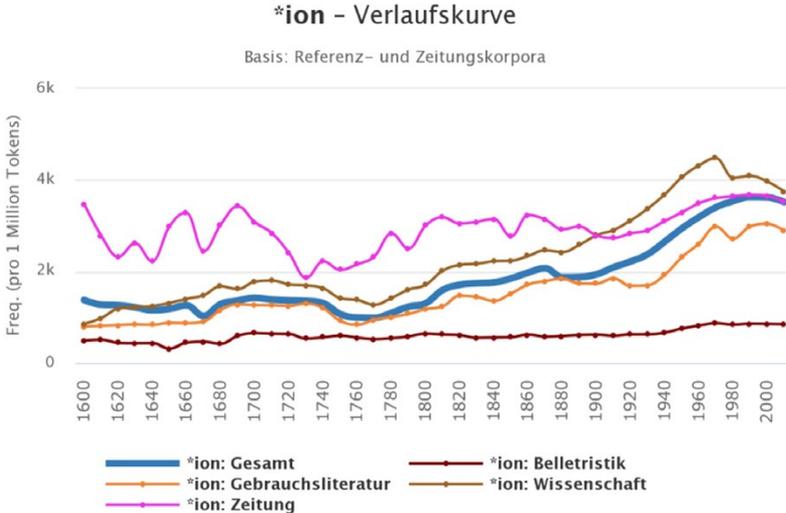


Abb. 1: ‚DWDS‘-Wortverlaufskurve für **ion* in den DWDS-RZ, sortiert nach Textsorten, erstellt durch das ‚DWDS‘ (Stand 03/2023)

Was die Derivate auf *-ierung* anbelangt, sind in den DWDS-RZ insgesamt 1 247 847 Tokens zu verzeichnen, die $f_{(r)}$ beläuft sich auf 936,1 pMW. Somit zeigt sich im Hinblick auf die Okkurrenz der Substantive auf *-ion* und *-ierung* in den DWDS-RZ eine ähnliche Proportionalität wie im DeReKo-W: Die Substantive auf *-ion* sind auch in diesem Korpus wesentlich häufiger belegt als die auf *-ierung*.

Neben den IDS-Korpora und den DWDS-Korpora können für korpusbasierte Analysen auch weitere Textsammlungen herangezogen werden, wie beispielsweise die Korpora des Projekts ‚Deutscher Wortschatz‘ der Universität Leipzig (vgl. URL 3). Das Projekt wurde von der Abteilung Automatische Sprachverarbeitung am Institut für Informatik der Universität Leipzig ins Leben gerufen und entwickelt (s. dazu Goldhahn/Eckart/Quasthoff [2012] u. Quasthoff/Richter [2005]). Es geht um ein Sprachportal, das zurzeit Zugriff auf mehr als 1 000 monolinguale korpusbasierte Wörterbücher in über 290 Sprachen ermöglicht (Stand 04/2023). Daneben bietet es verschiedene weitere Anwendungen wie Korpusstatistiken, diverse Analysewerkzeuge, darunter auch die Möglich-

keit, eigene Korpora aufzubereiten und zu verwalten usw. Im Rahmen dieses Projekts stehen verschiedene Nachrichtenkorpora zur Verfügung, wobei eines der umfangreichsten das Deutsche Nachrichten-Korpus (NewsCrawl-public 2018) (im Folgenden WUL-DN) mit 720 421 868 Tokens (Stand 04/2023) darstellt. Das Projekt bietet primär die Möglichkeit einer Recherche in einsprachigen Wörterbüchern, die Zugriff auf Beispiele aus verschiedenen Nachrichtenquellen ermöglicht. Allerdings wurden aus rechtlichen Gründen zahlreiche Metadaten über die Dokumente entfernt, so dass in diesen Wörterbüchern vor allem Beispielsätze ohne einen breiteren Kontext zu finden sind. Auf der anderen Seite bietet dieses Wörterbuch Auskunft über Kookkurrenzprofile entsprechender Ausdrücke einschließlich ihrer Synsemantika. Recherchen mit verschiedenen Suchanfragen sind in diesem Wörterbuch dennoch nicht möglich, so dass auch keine Angaben über die Gesamtzahl der Substantive auf *-ion* und *-ierung* gemacht werden können.

Aus den oben beschriebenen Korpusrecherchen geht hervor, dass sowohl die Substantive auf *-ion* als auch die auf *-ierung* im Lexikon des Deutschen eine sehr wichtige Rolle spielen, da sie eine hohe relative Frequenz aufweisen, und dass die ermittelten Frequenzen zum einen durch die Korpusgröße, zum anderen durch die Korpuszusammensetzung beeinflusst werden. Wie oben ausgeführt wurde, ist die Anzahl dieser Substantive jedoch zu hoch, damit einzelne Konkurrenzformen allein mithilfe der oben erwähnten Korpora identifiziert werden könnten. Einen ersten Überblick über die Anzahl der Parallelbildungen auf *-ierung* sowie *-ion* kann man sich anhand der Sekundärliteratur sowie anhand verfügbarer rückläufiger Wörterbücher verschaffen. Schäublin (1972), der sich im Rahmen seiner Analysen des adnominalen Attributs im Deutschen als einer der wenigen Autoren mit der Konkurrenz der Bildungen auf *-ion* und *-ierung* auseinandersetzte, sammelte insgesamt 138 konkurrierende Derivate wie *Artiklierung/Artikulation*, *Denunzierung/Denunziation* u. a., und zwar anhand von Exzerpten aus dem damals verfügbaren rückläufigen Wörterbuch von Mater aus dem Jahre 1965. Dabei schätzt Schäublin den Anteil an Parallelbildungen an allen deverbale Derivaten mit den Suffixen *-ung* und *-ion* auf ein Drittel ein. Er führt jedoch keine vollständige Liste an und bemerkt dazu, dass dies auch nicht das Ziel seiner Analyse ist (Schäublin 1972:64). Darüber hinaus nennt er noch weitere 31 Paare, die bei Mater (1965) nicht angeführt sind. In anderen Wörterbüchern ist die Anzahl dieser Doppelformen noch höher, beispielsweise sind in der elektronischen Version des ‚Duden. Fremdwörterbuchs‘ (2015) insgesamt 174 konkurrierende Derivate auf *-ierung* und *-ion*, im rückläufigen Wörterbuch von Mater (2001) sogar 255 hinterlegt.

Rückläufige Wörterbücher geben allerdings keine Antwort darauf, von welchen Verben auf *-ieren* entsprechende Derivate auf *-ierung* und/oder *-ion* im

Sprachgebrauch belegt sind und von welchen nicht, weshalb ich im Rahmen meiner Analyse anders vorging, und zwar ging ich von einer Datenbank von Verben auf *-ieren* aus, die aus ‚Duden, Das große Fremdwörterbuch‘ (2007) (im Folgenden ‚GFWD‘) exzerpiert wurden.⁵ In der ursprünglichen Datenbank waren insgesamt 3 898 Verben auf *-ieren* gebucht, wovon alle Verben aussortiert wurden, welche mit den stilistischen Markierungen ‚veraltet, veraltend, landschaftlich, regional, österreichisch, schweizerisch, scherzhaft, bayrisch‘ versehen sind, denn im Fokus meiner Analyse stand die deutsche Standardsprache.⁶ Insgesamt wurden im Rahmen meines Habilitationsprojekts 2 487 Verben auf *-ieren* im Hinblick auf die Okkurrenz entsprechender Derivate auf *-ierung* und *-ion* analysiert, wobei der Anteil an den Verbalisierern *-ier-*, *-isier-*, *-ifizier-* und *-izier-* folgendermaßen ausfällt:

Verben mit dem Verbalisierer <i>-ier-</i>	1.971 (etwa 79,3 %)
Verben mit dem Verbalisierer <i>-isier-</i>	434 (etwa 17,5 %)
Verben mit dem Verbalisierer <i>-ifizier-</i>	72 (etwa 2,9 %)
Verben mit dem Verbalisierer <i>-izier-</i>	10 (etwa 0,4 %)
Gesamtanzahl der untersuchten Verben auf <i>-ieren</i>	2.487 (100 %)

TAB. 2: Anzahl der untersuchten Verben auf *-ieren* mit den vier Suffixen *-ier-*, *-isier-*, *-ifizier-* und *-izier-*

Die Suche nach entsprechenden Formen erfolgte einerseits im DeReKo-W, andererseits wurde zu deren Identifizierung noch ein weiteres Vergleichskorpus herangezogen – das Google Books Korpus (Deutsch) (im Folgenden GBKD). Google Books ist eine sehr umfangreiche Textsammlung, die zurzeit mehr als 40 Mio. eingescannte Bücher in mehr als 500 Sprachen umfasst (s. URL 4), darunter auch auf Deutsch verfasste Bücher. Das an Google Bücher gekoppelte Korpustool Google Books NGram Viewer ermöglicht es, Verlaufskurven angegebener Ausdrücke entsprechend verschiedenen Suchanfragen abzurufen (zu Google Books NGram Viewer s. u. a. Brückner [2009, 2012], Rogelio/Renau [2012], Willems [2012]). Der Umfang des deutschen Korpus im Rahmen des NGram Viewers betrug nach Michel/Shen/Aiden (2011) etwa 36 Mrd. Tokens

-
- 5 Die modifizierte Datenbank ist auf der Basis einer Liste mit Verben auf *-ieren* entstanden, die mir freundlicherweise Carmen Scherer, die sich im Rahmen ihrer Forschung mit Argumentstrukturen entlehnter Verben befasst (s. Holler/Scherer 2010), zur Verfügung gestellt hat.
- 6 Allerdings wurden in der Datenbank diejenigen Verben auf *-ieren* beibehalten, bei denen mehrere Lesarten angegeben sind, worunter beispielsweise nur einige mit der stilistischen Markierung (österreich., schweizerisch) versehen sind.

(zit. Willems 2012:82). Heutzutage ist das Korpus deutlich größer, dennoch sind keine genauen Angaben über die aktuelle Größe des deutschsprachigen Korpus zu finden.

Eine Kombination mehrerer sehr großer Korpora mit unterschiedlicher Korpuszusammensetzung bezüglich verschiedener Textsorten und Kommunikationsbereiche erwies sich bei der Recherche als sehr nutzbringend, denn so konnten viel mehr Derivate identifiziert werden, als wenn nur das DeReKo herangezogen worden wäre. Zahlreiche Parallelbildungen auf *-ierung* und *-ion*, die allen voran im fachsprachlichen Kontext gebraucht werden, sind nämlich nur im GBKD enthalten, das im Vergleich mit dem DeReKo-W deutlich mehr Fachtexte umfasst. Um dies mit einigen Beispielen zu veranschaulichen: Rund 26,6 % der *-ung*-Derivate (z. B. *Disharmonierung*, *Glacierung*, *Hasardierung*, *Rezidivierung* u. v. a.) und rund 20,8 % der *-ion*-Derivate (z. B. *Dijudikation*, *Fistulation*, *Gargarisation*, *Suszeption* u. v. a.) sind lediglich im GBKD zu verzeichnen. Etwa bei 5 % der untersuchten Verben lassen sich beide Parallelbildungen antreffen, aber sie sind nur im GBKD belegt. Zu nennen sind etwa *Dialysierung/Dialysation*, *Hämolysierung/Hämolysation* aus der Fachsprache der Medizin, *Fluidifizierung/Fluidifikation*, *Gallierung/Gallation* aus der Fachsprache der Chemie, *Fistulierung/Fistulation*, *Resonierung/Resonation* aus der Fachsprache der Musik oder *Exkutierung/Exkution*, *Redhibierung/Redhibition* aus der Rechtssprache. Die Verlaufskurven des Google Books NGram Viewers zeigen, dass sich das proportionale Verhältnis der Konkurrenzformen im Laufe der Zeit dynamisch entwickelt und bei jedem solchen Paar individuell ist, wie an den folgenden Konkurrenzformen *Dialysierung/Dialysation* und *Fistulierung/Fistulation* zu sehen ist: Während die Konkurrenzformen *Fistulierung/Fistulation* im GBKD gleichmäßig verteilt sind und eine vergleichbare funktionale Auslastung aufweisen, macht sich am Derivat *Dialysation* bemerkbar, dass es allmählich durch die Konkurrenzform auf *-ierung* verdrängt wurde und dass das Derivat auf *-ierung* eine insgesamt höhere funktionale Auslastung aufweist (s. ABB. 2).

Aus der Analyse der Okkurrenz deverbaler Derivate im DeReKo-W geht zudem hervor, dass sich etwa einem Drittel aller untersuchten Verben (rund 33,2 %) sowohl Derivate auf *-ierung* als auch auf *-ion* zuordnen lassen, was mit der Annahme von Schäublin (1972) übereinstimmen würde. Wird neben dem DeReKo-W jedoch auch das GBKD berücksichtigt, lassen sich beide Parallelbildungen bei rund 63,8 % aller analysierten Verben identifizieren. Dies entspricht insgesamt 1 579 Parallelbildungen auf *-ion/-ierung* (etwa zwei Drittel aller untersuchten Verben), wobei dieser Befund darauf hindeutet, dass im Deutschen viel mehr Parallelbildungen bestehen, als in der Fachliteratur bisher angenommen wurde. Eine Kombination beider Korpora trägt somit zu wesentlich genaueren

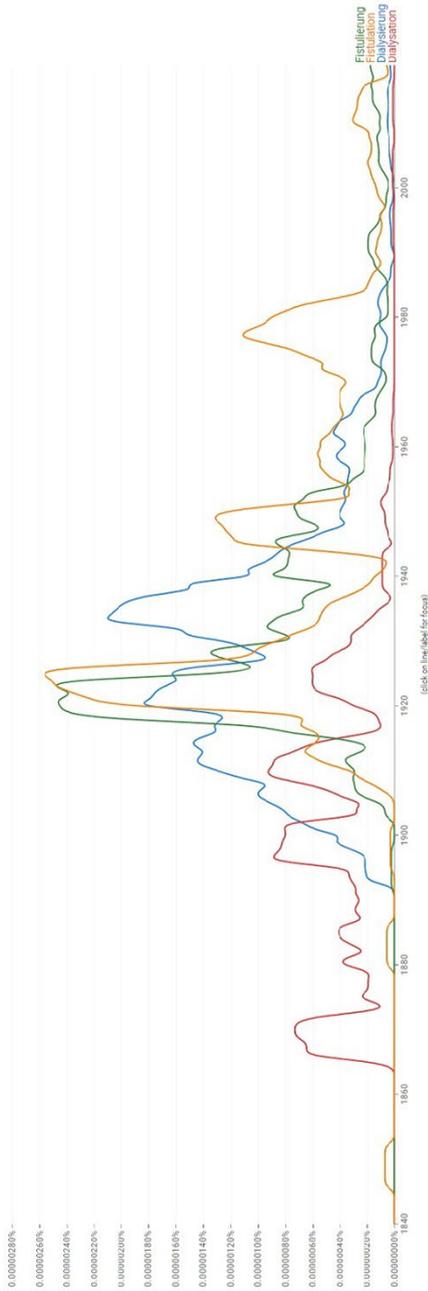


ABB. 2: Verlaufskurven von *Dialysierung/Dialyisation* und *Fistulierung/Fistulation* im GBKD (1840–2019), erstellt durch Google Books NGram Viewer, (Stand 04/2023)

Ergebnissen bei. Dank der Heranziehung beider sehr großen Korpora hat sich obendrein herausgestellt, dass von den 2 487 untersuchten Verben auf *-ieren* insgesamt 2 347 deverbale Derivate auf *-ierung* zu verzeichnen sind, was rund 94,4 % aller untersuchten Verben entspricht. Diese Feststellung steht im krassen Widerspruch zur Annahme von Kurth (vgl. 1956:307 ff.), der die Meinung vertritt, dass von den 1 500 Verben auf *-ieren*, die er untersuchte, lediglich ein Drittel Bildungen auf *-ierung* zulässt (Kurt 1956 zit. nach Schippan 1967:62). Auch Schäublin (1972:65) nennt eine Reihe von Verben auf *-ieren*, wie beispielsweise *abstrahieren*, *degenerieren*, *fungieren*, *gravitieren*, *konspirieren*, *meditieren*, *vegetieren*, *vibrieren* u. a., bei denen er kein Derivat auf *-ierung* präsupponiert. Im DeReKo-W liefert die Suche nach entsprechenden Derivaten mit dem Suffix *-ung* lediglich die Bildungen *Abstrahierung* ($f_{(a)} = 295$, $f_{(r)} = 0,0274$ pMW) und *Degenerierung* ($f_{(a)} = 144$, $f_{(r)} = 0,0134$ pMW), die anderen Derivate sind im DeReKo-W nicht hinterlegt (Stand 04/2023). Zieht man zur Recherche noch das GBKD heran, sind alle Derivate, d. h. auch *Fungierung*, *Gravitierung*, *Konspiration*, *Meditierung*, *Vegetierung* und *Vibrierung* zu finden. Zahlreiche Bildungen stellen dennoch auch im GBKD lediglich Randerscheinungen dar, aber einige Formen, wie z. B. *Meditierung* oder *Vegetierung*, liefern im ersten Fall über 40, im zweiten über 30 einsehbare Treffer. Die hohe Anzahl an Derivaten auf *-ierung* deutet darauf hin, dass das Wortbildungsmodell [exogener Stamm bzw. Konfix] + [Verbalisierer *-ier-* (i. w. S.)] + [natives Suffix *-ung*] im Deutschen (insbesondere in der Vergangenheit) eine sehr hohe Wortbildungsproduktivität aufweist und im Vergleich mit den *-ion*-Derivaten praktisch keinen morphologischen Bildungsbeschränkungen unterliegt. Dies korreliert damit, dass sich nur etwa 3 % der untersuchten Verben auf *-ieren* weder Derivate auf *-ierung* noch auf *-ion* zuordnen lassen (z. B. *abpassieren*, *albeszieren*, *belkantieren*, *effluieren*, *kratikulieren* u. a.) sowie damit, dass nur bei etwa 1 % der untersuchten Verben lediglich die Derivate auf *-ion* zu verzeichnen sind (z. B. *Dissertation*, *Deviation*, *Interaktion*).

Was die Verben auf *-ieren* anbelangt, denen sich in den untersuchten Korpora keine Derivate auf *-ion*, sondern nur auf *-ierung* zuordnen lassen, wurden insgesamt 743 Verben identifiziert (rund 30 % aller untersuchten Verben, davon sind 7,9 % lediglich im GBKD zu verzeichnen). Es handelt sich beispielsweise um Derivate wie *Arrangierung*, *Balancierung*, *Blamierung*, *Fusionierung*, *Möbliierung*, *Parodierung*, *Schockierung* u. a.). Der höhere Prozentanteil der Derivate auf *-ierung*, die keine Konkurrenzformen auf *-ion* zulassen, korreliert damit, dass sich hier zum Teil morphologische Bildungsbeschränkungen auswirken – z. B. sind keine *-ion*-Derivate von Verben auf *-ieren* möglich, deren Derivationsbasis ein *-ion*-Substantiv darstellt (**Fusion*[*at/it/t*]*ion*), – oder wenn die systemische Stelle im Lexikon durch ein anderes (meist entlehntes) Derivat mit oder ohne Suffix

besetzt ist (z. B. *Arrangement, Balance, Blamage, Parodie, Schock* usw.). Dank einer Recherche im GBKD konnten auch zahlreiche Derivate auf *-ion* identifiziert werden, die in anderen Sprachen, etwa dem Englischen oder Französischen anzutreffen, aber die in auf Deutsch verfassten Texten nicht belegt sind (etwa 10 % aller untersuchten Verben auf *-ieren*). Im Deutschen sind diese potenziell möglich, aber in keinem der untersuchten Korpora hinterlegt, wie beispielsweise *?Anonymisation, ?Astrologisation, ?Budgetation, ?Edition, ?Eternisation, ?Idylisation, ?Judaisation, ?Mythologisation, ?Tabuisation, ?Theologisation* u. v. a.

2.2. Quantitative Erfassung der Formen – Analyse der Frequenz

Nach der Identifizierung der Formen ist ihre funktionale Auslastung in verschiedenen Kontexten zu untersuchen. Hierzu sind quantitative Daten zu deren Okkurrenz in verschiedenen sehr großen Korpora notwendig, da diese möglichst viele Treffer liefern und genauere quantitative Analysen ihrer syntagmatischen Relationen ermöglichen. Die Frequenz stellt nämlich einen wichtigen Indikator der korpusbezogenen Usualität dar, wobei hier insbesondere die absolute, die relative und schließlich die proportionale Frequenz ($f_{(p)}$) erfasst werden. Bei quantitativen Analysen der semantischen Nähe und Distanz der zu analysierenden Ausdrücke gilt, dass sich die Aussagekraft der $f_{(p)}$ mit der steigenden $f_{(a)}$ erhöht. Bei der Ermittlung der Frequenz der Parallelbildungen können einige Problemfälle auftreten, auf die im Folgenden kurz eingegangen wird.

Manche Derivate auf *-ion* zeichnen sich durch die gleiche Lautung im Deutschen und zugleich in weiteren Sprachen, etwa dem Englischen oder dem Französischen, aus, was sich bei quantitativen Analysen als ein erheblicher Nachteil herausstellt. Dies macht sich beispielsweise bei Derivaten wie *Standardisation* bemerkbar. Die Suchanfrage *&Standardisation* liefert im DeReKo-W insgesamt 78 Treffer (Stand 04/2023), wenn man in den Suchoptionen kleingeschriebene Substantive aussortiert, 51 Treffer. Die meisten Belege entstammen jedoch englischsprachigen Textsegmenten, meistens handelt es sich um Bezeichnungen verschiedener Institutionen oder Normen wie *European Standardisation Committee* oder *International Organisation for Standardisation*, so dass die einzelnen Belege manuell durchgesehen werden müssen und ihre Anzahl deutlich nach unten zu korrigieren ist. Bei so einer hohen Anzahl an zu analysierenden Formen steigt der Arbeitsaufwand immens, auch wenn die Suchanfragen mithilfe trefferausschließender Operatoren wie z. B. *Standardisation %w3,s0* (*European* oder *Committee* oder *International*) etwas präzisiert werden können. Eine nachträgliche Aussortierung unpassender Treffer bringt einen beträchtlichen Zeitaufwand mit sich und ist im Falle von sehr hohen Frequenzen kaum realisierbar. In diesem Zusammenhang tritt ein weiteres Desiderat der künftigen Korpuslinguistik zu Tage, und zwar wäre eine Annotation fremdsprachiger Textsegmente

in auf Deutsch verfassten Texten nicht nur im DeReKo, sondern auch in den DWDS-Korpora und in den WUL-Korpora wünschenswert. Bei der automatischen Erkennung und Analyse von Textsegmenten nimmt in der Gegenwart die Bedeutung der Künstlichen Intelligenz enorm zu, sodass quantitative Analysen immer präziser werden und in Zukunft noch genauere Daten liefern werden.

Des Weiteren können bei der Ermittlung der Frequenz auch einige Probleme auftauchen, die sich auf die Ausführung konkreter Suchanfragen im Zusammenhang mit den *-ion*-Derivaten im DeReKo-W beziehen. Die Standardsuchanfrage mit dem Operator $\&$, mithilfe dessen alle Wortformen des entsprechenden Ausdrucks gesucht werden, liefert nicht selten keine Treffer, auch wenn Belege im analysierten Korpus doch zu verzeichnen sind. Dies ist insbesondere bei denjenigen Derivaten der Fall, die sich durch eine sehr niedrige Frequenz auszeichnen, wie etwa *Ätherisation* (die Suchanfrage $\&$ *Ätherisation* liefert keinen einzigen Treffer, die Suchanfrage *Ätherisation oder Ätherisationen* 2 Treffer), *Importation* (die Suchanfrage $\&$ *Importation* liefert keinen Treffer, die Suchanfrage *Importation oder Importationen* 17 Treffer, davon 16 fremdsprachig) oder *Charakterisation* (die Suchanfrage $\&$ *Charakterisation* liefert keinen einzigen Treffer, die Suchanfrage *Charakterisation oder Charakterisationen* 3 Treffer) (Stand 04/2023). Mitunter taucht dieses technische Problem jedoch auch bei Derivaten auf, die eine hohe Frequenz aufweisen, wie beispielsweise das Derivat *Signalisation* (s. ABB. 3).

Wird die entsprechende Suchanfrage jedoch *Signalisation oder Signalisationen* formuliert, ergibt die Suche fast 4 700 Treffer (s. ABB. 4).

Wie bereits oben erwähnt wurde, wurden bei den einzelnen Bildungen auf *-ierung* und *-ion* anhand ihres Vorkommens im DeReKo-W die $f_{(a)}$, $f_{(r)}$ (pro Million Wörter) und $f_{(p)}$ (in Prozent) errechnet, um Erkenntnisse über ihre funktionale Auslastung und bzw. auch über ihr Konkurrenzverhältnis zu erhalten. Die Frequenzwerte wurden in eine Datenbank eingetragen, die nach verschiedenen Kriterien sortiert werden kann (s. ABB. 5).

Im einleitenden Kapitel wurde bereits angedeutet, dass mit Nomina actionis auf *-ierung* sowie *-ion* bisweilen Infinitivkonversionen konkurrieren. Infinitivkonversionen wurden dennoch bei der Frequenzanalyse weder berücksichtigt noch in die Datenbank aufgenommen, da die Suchanfragen in morphosyntaktisch nicht annotierten Korpora nicht zwischen Form und Funktion von gesuchten Ausdrücken differenzieren können. Dementsprechend liefert die Suchanfrage $\&$ *Isolieren* sowohl Belege, in denen der gesuchte Ausdruck als Infinitivkonversion Verwendung findet, als auch solche, in denen er als Infinitiv in der Anfangsstellung im Rahmen einer diskursgrammatisch hervorgehobenen Position auftritt, wie im folgenden Beleg aus dem St. Galler Tagblatt, in dem die Infinitivkonversion im Kontext der Politik vorkommt:

IDS UNIVERSITÄT WÜRZBURG DEUTSCHE SPRACHE
COSMAS II - Abmeldung - Recherche - Optionen

Hilfe

W - öffentlich - alle öffentlichen Korpora des Archivs
W (mit Neuaquisitionen) [1]
Deutsches Referenzkorpus DeReKo-2023-I

Aktuelles Archiv: W - Archiv der geschriebenen Sprache
Aktuelle Suchanfrage: &Signalisation
Treffer: 0

Aktuelles Korpus: W - öffentlich - alle öffentlichen Korpora des Archivs
W (mit Neuaquisitionen) [1]
Referenz: Deutsches Referenzkorpus DeReKo-2023-I
Aktive Treffer:

Archiv Korpus Suchanfragen Wortform Ergebnisse Kook. KWIC Volltext Export

Eingabe: &signalisation

Weggelassener Verknüpfungoperator bedeutet: Wortabstand /w1 logisches 'ODER'
Beispiele: Wortformlisten:

Suchen

ABB. 3: Die Ergebnisse der Suchanfrage *&Signalisation* im DeReKo-W (Stand 04/2023)

IDS UNIVERSITÄT WÜRZBURG DEUTSCHE SPRACHE
COSMAS II - Abmeldung - Recherche - Optionen

Hilfe

W - öffentlich - alle öffentlichen Korpora des Archivs
W (mit Neuaquisitionen) [1]
Deutsches Referenzkorpus DeReKo-2023-I

Aktuelles Archiv: W - Archiv der geschriebenen Sprache
Aktuelle Suchanfrage: Signalisation oder Signalisationen
Treffer: 4.698

Aktuelles Korpus: W - öffentlich - alle öffentlichen Korpora des Archivs
W (mit Neuaquisitionen) [1]
Referenz: Deutsches Referenzkorpus DeReKo-2023-I
Aktive Treffer:

Archiv Korpus Such Wortformen Ergebnisse Kook. KWIC Volltext Export

2 Wortformlisten

Liste zu: Signalisation : 2 Wortformen, alphab. aufsteigend
Liste zu: Signalisationen : 1 Wortform, alphab. aufsteigend

Ergebnisse

Impressum | Datenschutz | Zitierhilfe | © 2003 - 2023 IDS Mannheim, COSMAS II web, Version 2.4.2

ABB. 4: Die Ergebnisse der Suchanfrage *Signalisation oder Signalisationen* im DeReKo-W (Stand 04/2023)

A	B	C	D	E	F	G	H	I	J	K
1 Verb auf -ieren globalisieren	Bedeutung laut dem „GFDW“ (2007) globalisieren <zu ...ieren>: 1. weltweit verbreiten, erdumfassend gestalten. 2. nicht ins Detail gehend, allgemein analysieren, beurteilen. 3. wirtschaftliche Strukturen, weltweit ausdehnen u. vertiefen (Wirtsch.).	Typ -ieren-Verb	Derivat auf -ierung	absolute Frequenz	relative Frequenz	proportionale Frequenz	Derivat auf -ion	absolute Frequenz	relative Frequenz	proportionale Frequenz
955	globalisieren	-ieren-Verb	Globalisierung	120978	12,2900	99,99	Globalisation	8	0,0008	0,01
956	glossieren	-ieren-Verb	Glossifizierung	2878	0,2920	98,53	Glossifikation	43	0,0044	1,47
957	godromisieren	-ieren-Verb	Glossierung	123	0,0120	100,00	Glossation	0	0,0000	0,00
958	godromisieren	-ieren-Verb	Godromierung	0	0,0000	0,00	Godromation	0	0,0000	0,00
959	goutieren	-ieren-Verb	Goutierung	10	0,0010	100,00	NB	0	0,0000	0,00
960	graduieren	-ieren-Verb	Gradierung	83	0,0090	27,93	Gradation	240	0,0240	72,07
961	grafitieren	-ieren-Verb	Graduierung	1421	0,1444	88,43	Graduation	186	0,0189	11,57
962	grafitisieren	-ieren-Verb	Grafitierung	1	0,0001	100,00	Grafitation	0	0,0000	0,00
963	grafitisieren	-ieren-Verb	Grafitisierung	0	0,0000	0,00	Grafitisation	0	0,0000	0,00

ABB. 5: Ausschnitt aus der Datenbank deverbaler Derivate von Verben auf -ieren (Stand 05/2021), NB = weder im DeReKo-W noch im GBKD belegt, ? – nur in fremdsprachigen Texten des GBKD belegt

- (1) *Totschweigen will ihn die Parteispitze. „Einige von euch haben sicher erwartet, dass ich was zu Oskar sage“, lockt SPD-Chef Franz Müntefering am vorigen Montag in Bonn seine rund 500 Zuhörer einer Wahlkampfveranstaltung. „Hab ich aber keine Lust zu“, feixt er gleich darauf ins urplötzlich hellwache Publikum. „Isolieren“ will Müntefering seinen Vorvorgänger.* (DeReKo-W: A97/JUN.12097 St. Galler Tagblatt, 30.6.1997, Ressort: RT-NAB (Abk.); Konkurs als Chance wahrgenommen)

Bei 2 487 Verben auf *-ieren* wäre das Vorhaben, auch dazugehörige Infinitivkonversionen einer genauen Frequenzanalyse zu unterziehen, schlicht nicht machbar, da die Belege anschließend manuell durchgesehen werden müssten. Infinitivkonversionen werden jedoch im Rahmen von Detailanalysen, in denen es um die Ermittlung kontextueller Unterschiede bei der Distribution einzelner Konkurrenzformen geht, mit einbezogen (s. im folgenden Abschnitt).

Bei Detailanalysen hat es sich obendrein als sinnvoll erwiesen, das proportionale Verhältnis der Varianten auf *-ierung* und *-ion* nicht nur im DeReKo-W, sondern auch in den weiteren zwei oben erwähnten Korpora – DWDS-RZ und WUL-DN – zu vergleichen, um potenzielle Unterschiede aufzudecken und zugleich die Aussagekraft der ermittelten Ergebnisse im Hinblick auf die funktionale Auslastung beider Parallelbildungen zu erhöhen. In der folgenden Tabelle werden zur Veranschaulichung die Frequenzwerte der Derivate *Installation* und *Installierung* angeführt:

	DeReKo-W	DWDS-RZ	WUL-DN
$f_{(a)}$ von <i>Installation</i>	144.867	11.147	9.046
Anzahl aller Tokens	10.764.331.064	1.333.888.521	720.421.868
$f_{(r)}$ pMW	13,46	8,3568	12,5565
$f_{(p)}$ in %	94	95,67	96,24
$f_{(a)}$ von <i>Installierung</i>	9.249	505	353
Anzahl aller Tokens	10.764.331.064	1.333.888.521	720.421.868
$f_{(r)}$ pMW	0,859	0,3786	0,4900
$f_{(p)}$ in %	6	4,33	3,76

TAB. 3: Absolute, relative und proportionale Frequenz der Derivate *Installation* und *Installierung* in den Korpora DeReKo-W, DWDS-RZ und WUL-DN (Stand 06/2023)

Aus der Tabelle ist ersichtlich, dass, obwohl die Korpusgröße der drei untersuchten Korpora sehr unterschiedlich ist und bei der ermittelten $f_{(a)}$ und $f_{(r)}$ zum Teil deutliche Unterschiede in Erscheinung treten, die Differenz bei der $f_{(p)}$ nicht so hoch ausfällt und lediglich im einstelligen Prozentbereich liegt. Somit zeigt sich,

dass die funktionale Auslastung des *-ion*-Derivats im Vergleich mit der Bildung auf *-ierung* in einem beträchtlichen Teil des öffentlichen Sprachgebrauchs deutlich höher ist. Um festzustellen, ob die Unterschiede in der ermittelten Frequenz in den verschiedenen untersuchten Korpora statistisch signifikant sind, empfiehlt sich die Durchführung eines Signifikanztests, wozu beispielsweise der Chi-Quadrat-Test herangezogen werden kann. Dieser weist eine hohe Sensitivität auf und lässt sich auch entsprechend verschiedenen Fragestellungen modifizieren (s. dazu Mostýn 2019b). Stellt man beispielsweise die $f_{(r)}$ des Derivats *Installation* im DeReKo-W und im WUL-DN in der obigen Tabelle gegenüber, sind die Werte auf den ersten Blick vergleichbar. Dennoch beträgt der χ^2 -Wert in diesem Fall aufgerundet auf zwei Dezimalstellen 40,96 (der kritische Wert liegt aufgerundet auf zwei Dezimalstellen bei 3,84), sodass dieser Unterschied statistisch höchst signifikant ist (die Wahrscheinlichkeit, dass diese Verteilung signifikant ist, liegt in diesem Fall bei 99,99 %). Die Berechnung des χ^2 -Wertes kann beispielsweise auch dann sinnvoll sein, wenn der Unterschied in der $f_{(p)}$ der Variante auf *-ion* und *-ierung* sehr gering ausfällt, was im Falle der Parallelbildungen *Installierung/Installation* allerdings nicht der Fall ist. Die ermittelten Frequenzen geben einen ersten Überblick über die Verteilung beider Varianten in verschiedenen Korpora sowie über ihre funktionale Auslastung, jedoch ohne Bezug auf konkrete Kontexte. Um Genaueres über die Distribution der Derivate auf *-ion* und *-ierung* sowie dazugehöriger Infinitivkonversion in verschiedenen Kontexten und Bedeutungen festzustellen, sind dennoch detailliertere Kookkurrenzanalysen vonnöten. Auf diese wird im folgenden Abschnitt am Beispiel der Parallelbildungen *Installation, Installierung* sowie deren Konkurrenzform *Installieren* näher eingegangen.

3. Analyse der Kookkurrenz deverbalder Derivate von Verben auf *-ieren*

Bevor Näheres über die syntagmatischen Relationen der oben genannten Wortbildungsprodukte herausgefunden werden kann, werfen wir einen Blick in das ‚DOW‘, um Auskunft über verschiedene Lesarten zu erhalten. Im Interpretament zum Derivat *Installation* werden vier verschiedene Sememe angegeben, 1a) ‚das Installieren‘ (ein Nomen actionis zu *installieren* ‚(eine technische Vorrichtung, Anlage o. Ä.) anschließen, einbauen, einrichten‘), 1b) ‚installierte technische Anlage (in ihrer Gesamtheit)‘, 2. ‚Amtseinführung, besonders Einsetzung in ein geistliches Amt‘ (schweizerisch, sonst veraltet) und 3. ‚von einem Künstler, einer Künstlerin im Raum eines Museums o. Ä. hergestelltes Arrangement mit verschiedenartigen Objekten, wie Schriften, Malereien, Plastiken, Fundstücken u. a., die so angeordnet werden, dass eine ganz spezielle Gestaltung des Raums entsteht‘ (Kunstwiss.). Im Interpretament zum Lemma *Installierung* wird in der

Bedeutungsparaphrase lediglich die Infinitivkonversion ‚das Installieren‘ angeführt. Den Bedeutungsparaphrasen lässt sich entnehmen, dass die Derivate im Kontext der Technik, der Kunst und der (geistlichen) Institutionen Verwendung finden, wobei die *-ion*-Bildung sowohl als Nomen actionis als auch als Nomen acti auftritt, während beim *-ung*-Derivat der Gebrauch als Nomen actionis anzunehmen ist.

Genauere Informationen über die Verwendungskontexte beider Derivate liefern jedoch erst detailliertere Kookkurrenzanalysen, die in den Korpora DeReKo-W, WUL-DN und mithilfe des ‚DWDS‘-Wortprofils, das seit Juni 2023 auf einer erweiterten Datenbasis mit etwa 6 Mrd. laufenden Wortformen beruht (s. ‚DWDS‘-Wortprofil in neuer Version, URL 5), durchgeführt werden. Dadurch kann man ein umfassendes Bild über den Gebrauch erhalten, wobei die Aufmerksamkeit insbesondere auf substantivische und adjektivische Kookkurrenzpartner gerichtet wird, denn diese treten oft als vorangestellte bzw. nachgestellte Attribute nominaler Kerne auf und sind daher zur Disambiguierung der Bedeutung besonders geeignet. Die Kookkurrenzanalyse im erstgenannten Korpus mit dem Fokus vier Wörter links und vier Wörter rechts (Standardeinstellung) macht deutlich, dass das Derivat *Installation* sehr häufig im Kontext der Kunst anzutreffen ist, wobei es sowohl als Nomen actionis als auch als Nomen acti hinterlegt ist. Unter den 30 Kookkurrenzpartnern mit dem höchsten LLR-Wert⁷ (also mit hoher statistischer Signifikanz) sind beispielsweise Substantive wie *Skulptur* ($f_{(a)} = 4\,720$), *Fotografie* (2 351), *Performance* (2 263), *Künstler* (2 152), *Bild* (1 387), *Malerei* (1 252), *Ausstellung* (1.193), *Gemälde* (842) u. a. sowie die Adjektive *interaktiv* (696), *raumgreifend* (605), *begehrbar* (641), *künstlerisch* (520), *multimedial* (389) vertreten. Weitere Substantive beziehen sich auf den Kontext der (Computer-)Technik (Lesarten 1a und 1b im ‚DOW‘, s. o.) wie etwa die Substantive *Anlage* (1 627), *Software* (893), *Wartung* (739), *Photo- bzw. Fotovoltaikanlage* (872), *Solaranlage* (731) sowie die Adjektive wie beispielsweise *technisch* (1 067) und *elektrisch* (1 082) (Stand 06/2023).

Auch das ‚DWDS‘-Wortprofil liefert Hinweise darauf, dass das Derivat *Installation* insbesondere in den Kontexten Kunst und (Computer-)Technik Verwendung findet, ähnlich wie im DeReKo-W tritt hier der Kontext der (geistlichen) Institutionen unter den häufigen Kookkurrenzen nicht in Erscheinung.

7 Die sog. Log-Likelihood-Ratio ist ein in der Korpuslinguistik bei Kookkurrenzanalysen sehr häufig angewandtes Assoziationsmaß zur Bewertung der statistischen Signifikanz des gemeinsamen Auftretens zweier Ausdrücke (s. u. a. Šemelík 2016:242; Perkuhn/Belica 2004; Kratochvilová 2011:194).



ABB. 6: Das ‚DWDS‘-Wortprofil für *Installation*, erstellt durch das ‚DWDS‘ (Stand 06/2023)

Zum Vergleich werden noch die am häufigsten zu verzeichnenden Kookkurrenzen des WUL-DN angeführt. Auch diese deuten auf die Kontexte Kunst und (Computer-)Technik hin:⁸

<p><i>Malerei</i> (1 098), <i>Fotografie</i> (885), <i>Künstlers</i> (811), <i>Künstlerin</i> (788), <i>Skulptur</i> (682), <i>Künstler</i> (673), <i>Wartung</i> (590), <i>Ausstellung</i> (451), <i>App</i> (418), <i>Performance</i> (414), <i>Zeichnung</i> (411), <i>Software</i> (378), <i>Apps</i> (347), <i>Anlage</i> (267), <i>Bildhauerei</i> (264), <i>Nutzer</i> (253), <i>begehbare</i> (240), <i>sehen</i> (239), <i>Solaranlagen</i> (222), <i>Updates</i> (217), <i>Kunst</i> (207), <i>Photovoltaikanlage</i> (205), <i>Raum</i> (203), <i>Video</i> (194), <i>Galerie</i> (192), <i>raumgreifende</i> (191), <i>Arbeiten</i> (191), <i>Anlagen</i> (189), <i>Inbetriebnahme</i> (188)</p>

TAB. 4: Die häufigsten Kookkurrenzpartner des Derivats *Installation* im WUL-DN (Stand 06/2023)

Unter den ersten 30 Kookkurrenzpartnern des Derivats *Installation* sind ebenfalls verschiedene Substantive zu verzeichnen, die vor allem auf den Kontext Technik hindeuten, häufig handelt es sich um Ausdrücke, die sich entsprechend den Kontextbelegen dem Sachbereich Bau-, Sicherheits-, Umwelt- und Verkehrs- bzw. Militärtechnik zuordnen lassen, wie z. B. *Anlage* ($f_{(a)}=136$), *System* (109), *Photo-/ Fotovoltaikanlage* (104), *Überwachungskamera* (38), *Alarmanlage*

8 Die ‚WUL‘-Kookkurrenzen werden hier ohne Synsemantika angeführt. Dabei werden vor allem häufig auftretende Kookkurrenten berücksichtigt, die auf verschiedene Kontexte hindeuten. Für die vollständige Kookkurrenzliste s. ‚WUL‘.

(30), *Ampel* (28), *Blindenleitsystem* (13), *Raketenabwehrsystem* (20), u. a. Was die adjektivischen Kookkurrenzpartner anbelangt, sind die Ausdrücke wie *neu* (650), *eigen* (118), *geplant* (76), *rasch* (36) wenig bedeutungsspezifisch, sie sind nicht nur im Kontext der Technik eingebettet, sondern sie treten auch im Kontext der Politik bzw. Organisation/Institution auf. Auf die zwei letztgenannten Kontexte deuten substantivische Kookkurrenzpartner wie *Betriebsrat* (48), *Ausschuss* (40), *Untersuchungsausschuss* (24), *Landesrechnungshof* (20), *Kindergartengruppe* (15), *Citymanager* (15) hin. Diese Kookkurrenzen machen deutlich, dass das Derivat *Installierung* auch in der Bedeutung ‚Einsetzung in ein Amt/eine Funktion‘ bzw. ‚Errichtung eines (politischen, wirtschaftlichen oder verwaltenden) Organs‘ gebraucht wird. Des Weiteren fällt ins Auge, dass der Kontext Kunst unter den ersten 30 Kookkurrenzen nicht zum Vorschein kommt.

Das ‚DWDS‘-Wortprofil des Derivats *Installierung* bietet im Vergleich zum Wortprofil der Parallelbildung *Installation* nicht viele Kookkurrenzen:



ABB. 7: Das ‚DWDS‘-Wortprofil für *Installierung*, erstellt durch das ‚DWDS‘ (Stand 06/2023)

Die substantivischen Partner *Kamera*, *System* und *Raketenabwehrsystem* deuten auf den Gebrauch im Kontext der Technik i. w. S. hin, während die Kookkurrenzen *Regime* und *Übergangsregierung* auf Politik Bezug nehmen. Hier macht sich zugleich eine unterschiedliche Stärke der Assoziationen bemerkbar, denn während beispielsweise das Substantiv *Übergangsregierung* im ‚DWDS‘-Wortprofil zu einigen der stärksten Kookkurrenten gehört, nimmt dieser Kookkurrenzpartner im DeReKo-W erst den 106. Platz ein. Dies ist nicht nur durch die zum Teil unterschiedliche Zusammensetzung der Korpora bedingt, sondern es hängt wahrscheinlich auch damit zusammen, dass die Assoziationsmaße auf unterschiedlichen Methoden beruhen. Im DeReKo-W wird die Stärke der Kookkurrenzen, wie bereits oben erwähnt wurde, aufgrund der LLR-Ratio berechnet, während sie im ‚DWDS‘-Wortprofil auf dem logDice-Koeffizienten basiert. Beide Assoziationsmaße fußen zwar auf ähnlichen Prinzipien, doch sind die Berechnungsmethoden und die statistischen Grundlagen unterschiedlich. Daher ist es vorteilhaft, Kookkurrenzanalysen anhand von mehreren Korpora durchzuführen. Dies bestätigen auch die Ergebnisse der Kookkurrenzanalyse im

WUL-DN. Die automatisch generierten Kookkurrenzen zeigen nämlich, dass vom Derivat *Installierung* nicht nur in den bereits erwähnten Kontexten Technik und Politik bzw. Organisation/Institution Gebrauch gemacht wird, sondern auch in weiteren Kontexten wie etwa Sport. Darauf deuten die Kookkurrenzen wie beispielsweise *Frauen-Teams*, *FA* (= Football Association), *Fußballverband*, *Manchester (United)* hin:

neuen (53), *Informationssysteme* (47), *Frauen-Teams* (45), *Mautsysteme* (43), *Ansuchen* (35), *Systeme* (31), *FA* (31), *Polen* (30), *Photovoltaikanlage* (29), *Fußballverband* (26), *Tschechien* (25), *Firma* (24), *Schaffung* (24), *Bundesland* (24), *Misstrauen* (20), *Ministerien* (20), *Software* (20), *Manchester United* (19), *geplante* (18), *unabhängigen* (18), *modernen* (17), *Kameras* (16), *Mitgliederversammlung* (15), *Museen* (15), *United* (15), *Landeshauptmann* (15), *vorgesehen* (15), *laufende* (14), *Manchester* (14), *englischen* (14)

TAB. 5: Die häufigsten Kookkurrenzpartner des Derivats *Installierung* im WUL-DN (Stand 06/2023)

Im Rahmen der Analyse des Gebrauchs von Parallelbildungen auf *-ierung* und *-ion* ist noch ein weiteres nützliches Korpustool zu erwähnen, und zwar die Kookkurrenzdatenbank CCDB von Cyril Belica, die eine optimale Ergänzung zu den oben erwähnten Korpusinstrumenten darstellt. Es handelt sich um eine korpuslinguistische Denk- und Experimentierplattform, die bereits seit 2001 besteht und auf simulierten neuronalen Netzen beruht, wobei sie selbstorganisierende Karten – sog. Kohonennetze – generieren kann. Sie wurde auf der Grundlage eines auf dem DeReKo basierenden virtuellen Korpus geschriebener Gegenwartssprache des IDS im Umfang von etwa 2,2 Mrd. Wortformen entwickelt und ermöglicht es, Kookkurrenzprofile von mehr als 220 000 Lemmata abzurufen und ihre typischen syntagmatischen Muster zu untersuchen (vgl. Belica 2005, Keibel/Belica 2007, Belica et al. 2009, Belica 2001–2023; Heringer 2012:35; Marková 2012:42 ff.). CCDB bietet insgesamt fünf verschiedene Funktionen im Rahmen der Modellierung semantischer Nähe an. Zum Zweck der Analyse der Distribution der Parallelbildungen auf *-ierung* und *-ion* in verschiedenen Kontexten kann insbesondere die Funktion Kontrastierung naher Synonyme (contrast near synonyms) genutzt werden, mit deren Hilfe farbige CNS-Karten erstellt werden, die Aufschluss über Kontexte geben, in denen beide Ausdrücke gemeinsam auftreten und vornehmlich über Kontexte, in denen entweder das Derivat auf *-ierung* oder das auf *-ion* häufiger belegt ist.

Zu den Kontexten, in denen beide Parallelbildungen auftreten, gehören laut der folgenden CNS-Karte für die Derivate *Installation/Installierung* insbesondere Technik i. w. S., wie die Cluster in mittelgrauen Tönen (in der Originalfarbe

orange) wie A4–A5, B4–B5, C4–C5, D4–D5 und E5 mit den Ausdrücken wie etwa *Anlage, Einbau, Heizung, Konfigurationstechnik, Sanierung, Software* und *Zebrastrreifen* demonstrieren:

© Cyril Belica: Modelling Semantic Proximity - Contrasting Near-Synonyms (version: 0.21)

Installation : Installierung

export SOM as WMF or SVG file

Collage Zeichnung Siebdruck großformatig Aquarell Mischtechnik Radierung kleinformatig	Grafik skulptural Tafelbild zweidimensional kinetisch minimalistisch figürlich	dreidimensional visualisieren interaktiv grafisch multimedial audiovisuell Animation digital	Konfiguration Programmierung Software benutzerfreundlich Apparatur Anwender Bedienung Implementierung	Wartung Dämmung Telefonanlage Lüftung Elektroninstallation Verfälschung veraltet Belüftung
Skulptur Druckgrafik Plastik Fotoarbeit Druckgraphik Graphik Gemälde Papierarbeit	Videoinstallation Rauminstallation Fotografie Videoarbeit Malerei Videokunst konzeptuell Medienkunst	Performance Computeranimation Video Videoprojektion computergeneriert Computergrafik Kunstform Ausschnitt	Montage Steuerung bestücken	installieren installiert Heizung Einbau Warmwasserbereitung Anlage Heizungsanlage Heizanlage
Ausstellung Einzelausstellung Werkgruppe Schau Exponat ausgestellt Werk ausstellen	Klanginstallation Lichtinstallation Happening Konzeptkünstler Vernissage Gruppenausstellung Bourgeois Rist	Bure	ausstatten ausrüsten angeschafft tragbar schalten vollautomatisch Bewegungsmelder ausgerüstet	hochmodern Radargerät Apparat modern vollelektronisch Gerät Ausrüstung computergestützt
		Abschaltung Verlegung Einführung Entsendung Abberufung Aufklärung Freilassung Wiederbesetzung	Stationierung Überwachung Anbringung Raketenabwehr Erfassung Geschwindigkeitskontrol Umstellung Bereitstellung	Anschaffung Aufrüstung Nachrüstung Umrüstung Großanlage Neuanschaffung Zebrastrreifen Ankauf
Einsetzung Einberufung Sitzung Konstituierung Empfehlung Vorsitzender Ausländerfrage Vorsitzende	Gründung Einschaltung weilers Unterrichtung fernet außerdem Bildung Bürgerbefragung	Aufstockung Neubesetzung Neustrukturierung Ausweitung Überprüfung Reorganisation Abklärung Überarbeitung	Aufbau Schaffung Einrichtung Etablierung Funktionieren Erneuerung Informationskampagne Aufrechterhaltung	Errichtung Inbetriebnahme Bau Realisierung Erweiterung Ausbau Modernisierung Sanierung

ABB. 8: Kontrastierung von Beinahe-Synonymen *Installierung* und *Installation* anhand der Korpusplattform CCDB von Belica (Belica 1995, 2001–2023)

Um weitere Daten über die Distribution beider Parallelbildungen zu erhalten, erachte ich es als sinnvoll, eine anschließende Kookkurrenzanalyse im DeReKo-W mit ausgewählten signifikanten Kookkurrenzpartnern durchzuführen, die gleichzeitig bei beiden Derivaten als Kookkurrenzen vorkommen, wie in der folgenden Tabelle am Beispiel der Kontexte Bau- und Computertechnik demonstriert wird. Dabei wird auch das Konkurrenzverhältnis zur Infinitivkonversion *Installieren* unter die Lupe genommen. Die Suchanfrage mit der Infinitivkonversion wird

so formuliert, dass die Personalpronomen *Sie*, *sie* und *wir* im Kommando ausgeschlossen werden, womit zumindest gleichlautende finite Verbformen in der Anfangsstellung automatisch aussortiert werden, bevor sie manuell durchgesehen werden. Die Suchanfragen mit dem Wortabstand +w3 zielen darauf ab, u. a. Belege mit nachgestellten Genitiv- und Präpositionalattributen zu bekommen.

Suchanfrage im DeReKo-W	f _(a)	f _(r)	f _(p)
&Installation /+w3,s0 (&Programm oder Software oder &Treiber)	951	0,0883 pMW	86,93 %
&Installierung /+w3,s0 (&Programm oder Software oder &Treiber)	33	0,0031 pMW	3,02 %
(Installieren %+w1,s0 (Sie oder sie oder wir)) /+w3,s0 (&Programm oder Software oder &Treiber)	110	0,0102 pMW	10,05 %
Summe	1.094		100 %
&Installation /+w3,s0 (&Anlage oder &Beleuchtung oder &Heizung)	1.371	0,1274 pMW	89,08 %
&Installierung /+w3,s0 (&Anlage oder &Beleuchtung oder &Heizung)	133	0,0124 pMW	8,64 %
(Installieren %+w1,s0 (Sie oder sie oder wir)) /+w3,s0 (&Anlage oder &Beleuchtung oder &Heizung)	35	0,0032 pMW	2,27 %
Summe	1.539		100 %

TAB. 6: Verteilung der Konkurrenzformen *Installation*, *Installierung*, *Installieren* mit den Kookkurrenten *Programm*, *Software*, *Treiber*, *Anlage*, *Beleuchtung*, *Heizung* im DeReKo-W (Stand 06/2023)

Aus der Kookkurrenzanalyse ergibt sich, dass im Kontext der Computertechnik mit den in der obigen Tabelle angeführten Kookkurrenten *Programm*, *Software* und *Treiber* eindeutig das Derivat auf *-ion* überwiegt, wobei sich seine f_(p) auf fast 87 % beläuft. Die Infinitivkonversion, die als Konkurrenzform aufgrund zahlreicher morphologischer, syntaktischer und semantischer Beschränkungen in den meisten Fällen eine marginale Rolle spielt (s. dazu eingehend in Mostýn 2022b), ist im DeReKo-W im Kontext der Computertechnik dennoch häufiger als das Derivat auf *-ierung* belegt. Auch im Kontext der Bautechnik spielt das Derivat auf *-ion* eine dominante Rolle, was die Kookkurrenzen mit den substantivischen Partnern *Anlage*, *Beleuchtung* und *Heizung* zeigen. Dabei beträgt die f_(p) von *Installation* fast 90 %. Dagegen weist die Konkurrenzform auf *-ierung* in diesem Kontext eine wesentlich niedrigere funktionale Auslastung auf, was ebenfalls auf die Infinitivkonversion *Installieren* zutrifft. Vergleicht man diese

Ergebnisse mit den in der CNS-Karte, fällt auf, dass der vom CCDB-Tool angegebene Anteil *-ion:-ierung* in den Clustern wie etwa B5 mehr zugunsten der *-ierung*-Bildung ausfällt (66:34). Dies ist auf die unterschiedliche Größe der untersuchten Korpora sowie auf unterschiedliche Berechnungsmethoden zurückzuführen und unterstreicht noch einmal die Wichtigkeit einer Kombination mehrerer Tools und Korpora, um genauere Daten zu erhalten. Schließlich sollte noch die $f_{(r)}$ der Belege in Texten bundesdeutscher, österreichischer und Schweizer Provenienz einem Vergleich unterzogen werden, um potenzielle Unterschiede im Hinblick auf die areale Distribution aufzudecken. Es zeigt sich nämlich, dass einige Derivate auf *-ion* in der Schweiz eine deutlich höhere $f_{(r)}$ als in Texten bundesdeutscher oder österreichischer Herkunft aufweisen (s. dazu ebd.)

In der CNS-Karte für *Installierung* – *Installation* ist außerdem eine Reihe von hellgrau gefärbten Clustern zu sehen (A1–C3) (in der Originalfarbe Gelb), die substantivische Ausdrücke wie beispielsweise *Collage*, *Zeichnung*, *Grafik*, *Gemälde*, *Kunstform*, *Ausstellung*, *Exponat* sowie adjektivische Ausdrücke wie *skulptural*, *multimedial*, *audiovisuell*, *interaktiv* und einige Verben wie etwa *ausstellen* beinhalten. Die Ausdrücke in diesen Clustern weisen laut dieser CNS-Karte eine 100%ige Ähnlichkeit mit dem Kookkurrenzprofil von *Installation* in der dritten Lesart (s. o.) auf und deuten darauf hin, dass auch im Kontext der Kunst eindeutig das *-ion*-Derivat dominiert. Dies bestätigen auch detailliertere Kookkurrenzanalysen im DeReKo-W.

Im Gegensatz dazu stehen die dunkelgrauen Cluster D3 und E1–E3 (in der Originalfarbe Rot), in denen Ausdrücke wie *Verlegung*, *Entsendung*, *Abberufung*, *Einsetzung*, *Vorsitzende(r)*, *Neubesetzung*, *Reorganisation* u. a. zu finden sind, die auf den Kontext Politik bzw. Organisation/Institution bezogen sind und auf die Lesart ‚Einsetzung in ein Amt/eine Funktion‘ bzw. ‚Errichtung eines (politischen, wirtschaftlichen oder verwaltenden) Organs‘ und damit einhergehende strukturelle Veränderungen hindeuten. Die Kookkurrenzprofile der in diesen Clustern auftretenden Wörter ähneln dem Kookkurrenzprofil des Derivats *Installierung* und nicht dem von *Installation*, wie man ausgehend von den im ‚DOW‘ angeführten Bedeutungsparaphrasen annehmen würde. Das vom CCDB-Tool auf der Webseite angegebene Verhältnis *-ion:-ierung* beträgt beispielsweise 0:100 für D3 und E1 bzw. 1:99 für D2, was deutlich macht, dass im Kontext Politik bzw. Organisation/Institution das Derivat auf *-ierung* eindeutig vorherrscht.⁹ Um dies zu überprüfen, werden im Folgenden Kookkurrenzen mit

9 Ein vergleichbares Tool zur Kontrastierung des Gebrauchs von zwei Ausdrücken bietet auch das ‚DWDS‘-Profil, indem es den Nutzern ermöglicht, Kookkurrenzprofile von zwei Ausdrücken zu vergleichen und optional nach Unterschieden oder nach Gemeinsamkeiten ihrer syntagmatischen Muster zu suchen. Die Anzahl der Treffer für *Installierung* ist in diesem Korpus jedoch zu klein, um aussagekräftige Ergebnisse zu bekommen, so dass dieses Tool in diesem Fall dazu nicht geeignet ist. Daher wird wieder das DeReKo-W herangezogen.

den substantivischen Partnern *Arbeitsgruppe*, *Ausschuss*, *Betriebsrat* und *Regierung* untersucht, die auf verschiedene Organe oder Institutionen im Bereich der Wirtschaft, Verwaltung oder Politik Bezug nehmen, sowie Kookkurrenzen mit den Personenbezeichnungen *Beauftragter*, *Geschäftsführer*, *Manager* und *Sportdirektor*. Darüber hinaus wird der Fokus noch auf den Kontext des geistlichen Amtes gerichtet. Unter den signifikanten Kookkurrenzen sind im DeReKo-W allerdings nur sehr wenige Vertreter zu finden, sodass lediglich der substantivische Partner *Pfarrer* herangezogen wird.

Suchanfrage im DeReKo-W	f _(a)	f _(r)	f _(p)
&Installation /+w3,s0 (&Arbeitsgruppe oder &Ausschuss oder &Betriebsrat oder &Regierung)	53	0,0049 pMW	24,48 %
&Installierung /+w3,s0 (&Arbeitsgruppe oder &Ausschuss oder &Betriebsrat oder &Regierung)	154	0,0143 pMW	74,03 %
(Installieren %+w1,s0 (Sie oder sie oder wir)) /+w3,s0 (&Arbeitsgruppe oder &Ausschuss oder &Betriebsrat oder &Regierung)	1	0,0001 pMW	1,49 %
Summe	208		100 %
&Installation /+w3,s0 (&Beauftragter oder &Geschäftsführer oder &Manager oder &Sportdirektor)	42	0,0039 pMW	39,25 %
&Installierung /+w3,s0 (&Beauftragter oder &Geschäftsführer oder &Manager oder &Sportdirektor)	65	0,0060 pMW	60,75 %
(Installieren %+w1,s0 (Sie oder sie oder wir)) /+w3,s0 (&Beauftragter oder &Geschäftsführer oder &Manager oder &Sportdirektor)	0	0,0000 pMW	0,00 %
Summe	107		100 %
&Installation /+w3,s0 (&Pfarrer)	67	0,0062 pMW	81,71 %
&Installierung /+w3,s0 (&Pfarrer)	15	0,0014 pMW	18,29 %
(Installieren %+w1,s0 (Sie oder sie oder wir)) /+w3,s0 (&Pfarrer)	0	0,0000 pMW	0,00 %
Summe	82		

TAB. 7: Verteilung der Konkurrenzformen *Installation*, *Installierung*, *Installieren* mit den Kookkurrenten *Arbeitsgruppe*, *Ausschuss*, *Betriebsrat*, *Regierung*, *Beauftragter*, *Geschäftsführer*, *Manager*, *Sportdirektor* und *Pfarrer* im DeReKo-W (Stand 06/2023)

Aus der Kookkurrenzanalyse geht hervor, dass das Derivat *Installierung* im syntaktischen Umfeld der Kookkurrenzpartner *Arbeitsgruppe*, *Ausschuss*, *Betriebsrat* und *Regierung* deutlich vorherrscht. In fast drei Vierteln der Belege ist im DeReKo-W die *-ung*-Bildung belegt, während das *-ion*-Derivat in etwa einem

Viertel der Fälle gebraucht wird. Hier macht sich ein Unterschied im Hinblick auf das vom CCDB-Tool ermittelte Verhältnis *-ion:-ierung* bemerkbar, denn laut dem CCDB-Tool würde man in diesem Kontext eine deutlich niedrigere funktionale Auslastung des *-ion*-Derivats erwarten (s. o.). Im folgenden Beispiel kommt *Installierung* im Kontext der Etablierung einer demokratischen Regierung in Kambodscha in den 1990er Jahren vor, die durch die UN initiiert wurde.

- (2) *Die jüngste UN-Wahlmission in Asien ging schief und kostete zwei Milliarden Dollar. Truppen, darunter auch deutsche Medizinerteams und französische Fremdenlegionäre, hatten die Wahlen in Kambodscha „überwacht“. Die **Installierung** einer „demokratischen Regierung“ war als UN-Erfolg gefeiert worden.* (DeReKo-W: N99/SEP.36777 Salzburger Nachrichten, 3. 9. 1999, Ressort: Weltpolitik; Ost-Timor: Ein Meer von Feuer)

Das *-ion*-Derivat ist im obigen Beleg ohne Bedeutungsunterschied einsetzbar, der Substitutionstest ergibt den Satz (2a) *Die **Installation** einer „demokratischen Regierung“ war als UN-Erfolg gefeiert worden.* Die Infinitivkonversion kommt in diesem Satz als Ersatzform nicht in Frage, was nicht nur durch die syntagmatischen Beziehungen bedingt ist – die Infinitivkonversion *Installieren* ist nämlich in keinem der untersuchten Korpora als Kookkurrenzpartner von *Regierung* hinterlegt –, sondern auch durch die Temporalität, da Infinitivkonversionen das Geschehen oder die Handlung als ein Kontinuum zum Ausdruck bringen. Im obigen Beleg trägt die Ereignisnominalisierung *Installierung* jedoch eine resultative Komponente. Des Weiteren ist *Installieren* in der syntaktischen Funktion eines Subjekts nur äußerst selten zu verzeichnen, die Infinitivkonversion kommt in den untersuchten Korpora z. B. in Phrasen wie *das Installieren ist/war einfach/möglich* vor, wobei es sich um den Kontext der Computertechnik handelt. Infinitivkonversionen heben die Tätigkeit hervor, den einzigen Treffer mit den oben genannten Kookkurrenten stellt im DeReKo-W der Satz (3) *Ergebnisse sind aber wichtiger als das **Installieren** von Arbeitsgruppen.* (DeReKo-W: PBY/W14.00016 Protokoll der Sitzung des Parlaments Bayerischer Landtag am 21. 4. 1999. 16. Sitzung der 14. Wahlperiode 1998–2003. Plenarprotokoll, München, 1999). Dies deutet darauf hin, dass die Infinitivkonversion *Installieren* in der Bedeutung ‚Errichtung eines (politischen, wirtschaftlichen oder verwaltenden) Organs‘ nicht gebräuchlich zu sein scheint. Dies betrifft auch die Lesart ‚Einsetzung in ein Amt/eine Funktion‘. Die Distribution in Syntagmen mit den Personenbezeichnungen *Beauftragter*, *Geschäftsführer*, *Manager* und *Sportdirektor* machen das Vorherrschen des *-ung*-Derivats deutlich, wobei seine $f_{(p)}$ rund 61 % beträgt, die des *-ion*-Derivats beläuft sich auf etwa 39 %. Dass die Infinitivkonversion in Kookkurrenzen mit diesen Substantiven nicht anzutreffen ist, offenbart auch die Suche im GBKD. Diese wird vor allem dann

ergänzend zur Recherche im DeReKo-W durchgeführt, wenn es im DeReKo-W um sehr seltene Belege geht.

Im DeReKo-W sind Belege, in denen die Bildungen *Installation*, *Installierung* in der Lesart ‚eine Amtseinführung in ein geistliches Amt‘ Verwendung finden, nicht häufig zu verzeichnen. Hier konkurrieren lediglich die Derivate *Installation* und *Installierung*. Die in der obigen Tabelle angeführte Suchanfrage mit dem Kookkurrenzpartner *Pfarrer* hat insgesamt 82 Treffer geliefert, wobei die *-ion*-Bildung mit einer $f_{(p)}$ im Wert von fast 82 % eindeutig dominiert. Die $f_{(p)}$ von *Installierung* beträgt nur etwa 18 %, die Infinitivkonversion ist in Syntagmen mit dem Kookkurrenten *Pfarrer* im DeReKo-W nicht zu verzeichnen.

4. Fazit

Ausgehend von den hier präsentierten Überlegungen zur korpuslinguistischen Analyse der Derivate auf *-ierung* und *-ion* lässt sich festhalten, dass die heutige Korpuslinguistik mit verschiedenen Korpora und Korpusinstrumenten zahlreiche Möglichkeiten bietet, die Form und Funktion von synonymen Ausdrücken näher zu untersuchen. Analysen syntagmatischer Relationen mithilfe von Kookkurrenzanalysen geben Aufschluss über ihre Distribution in verschiedenen Kontexten und sind bei der Disambiguierung ihrer Bedeutung und ihres Gebrauchs sehr hilfreich. Bevor detaillierte Kookkurrenzanalysen durchgeführt werden können, ist es zunächst wichtig, die einzelnen Derivate auf *-ierung* und *-ion* zu identifizieren. Zu diesem Zweck sind sehr große Korpora wie das DeReKo heranzuziehen. Aus dem Vergleich der vier Archive der geschriebenen Sprache des DeReKo ergibt sich dennoch, dass bei der Korpusauswahl nicht nur die Korpusgröße, sondern auch die Korpuszusammensetzung zu berücksichtigen sind. Darüber hinaus hat die Analyse ergeben, dass zahlreiche Parallelbildungen auf *-ierung* und *-ion* lediglich im GBKD belegt sind, weil sie oft im fachsprachlichen Kontext gebraucht werden, der im GBKD stärker vertreten ist. Dies weist auf einen höheren Fachlichkeitsgrad dieses Korpus hin. Eine Kombination mehrerer sehr großer Korpora wie des DeReKo-W und des GBKD trägt also wesentlich dazu bei, die Aussagekraft von Korpusanalysen zu erhöhen, indem viel mehr Parallelbildungen identifiziert werden konnten, als wenn nur das DeReKo-W untersucht worden wäre. Die Aufbereitung eines umfangreichen Fachsprachenkorpus mit fortgeschrittenen Suchfunktionen bleibt nichtsdestotrotz weiterhin ein großes Desiderat der gegenwärtigen Korpuslinguistik, denn die Suchfunktionen im GBKD und insbesondere die Möglichkeiten genauerer quantitativer Analysen sind in diesem Korpus sehr begrenzt.

Nach der Identifizierung der Formen anhand des DeReKo-W und des GBKD soll eine quantitative Erfassung ihrer Frequenz erfolgen, insbesondere

die Ermittlung der absoluten, der relativen und der proportionalen Frequenz, denn die Frequenz stellt einen wichtigen Anzeiger der korpusbezogenen Usualität dar. Bei der quantitativen Analyse sind einige Problemfälle aufgetaucht, die sich im Bereich der Derivate auf *-ion* insbesondere auf die morphosyntaktische Annotation fremdsprachiger Textsegmente und ihren Status als Wortbildungsprodukt oder auf die Form und Funktion von Infinitiven und dazugehörigen Infinitivkonversionen beziehen. Da die Möglichkeiten morphosyntaktischer und auch semantischer Annotationen mithilfe der Künstlichen Intelligenz immer vorangetrieben werden, werden quantitative Analysen in Korpora in Zukunft hoffentlich einfacher durchzuführen sein und verlässlichere Daten liefern. Bei der quantitativen Analyse der Derivate auf *-ierung* und *-ion* ist es sinnvoll, verschiedene große Korpora miteinander zu vergleichen, um die Validität im Hinblick auf die ermittelten Frequenzen beider Parallelbildungen zu sichern. Als nutzbringend haben sich bei der Analyse neben dem DeReKo-W, aufgrund dessen eine Datenbank deverbaler Derivate mit 2 487 Einträgen entstanden ist, auch die Korpora DWDS-RZ und WUL-DN erwiesen, die insbesondere bei detaillierteren Analysen der Okkurrenz ausgewählter Derivate und ihrer Ko-okkurrenz ein umfassendes Bild ihres Gebrauchs bieten und einander ergänzen können, wie am Beispiel der Parallelbildungen *Installation* und *Installierung* demonstriert wurde. Um festzustellen, ob die ermittelten Frequenzen statistisch signifikant sind, empfiehlt es sich, einen Signifikanztest durchzuführen, wozu beispielsweise der Chi-Quadrat-Test herangezogen werden kann.

Bevor eingehende Kookkurrenzanalysen durchgeführt werden, die Aufschluss über die funktionale Auslastung der Konkurrenzformen in verschiedenen Kontexten geben, sollten noch entsprechende Interpretamente (falls vorhanden) in einsprachigen Wörterbüchern zu Rate gezogen werden, um verschiedene Lesarten der Parallelbildungen zu vergleichen und Kontexte zu identifizieren, in denen sie auftreten. Ein methodisches Vorgehen zur anschließenden Kookkurrenzanalyse ließe sich in drei Schritten wie folgt zusammenfassen:

- I. Kookkurrenzanalyse I – Analyse dominanter syntagmatischer Muster beider Parallelbildungen (Kookkurrenzprofile anhand verschiedener Korpora und Korpusinstrumente DeReKo-W – Cosmas II, DWDS – Wortprofile, Wortschatz-Portal der Universität Leipzig – Integriertes Korpusool zur Kookkurrenzanalyse)

Eine Analyse dominanter syntagmatischer Muster beider Parallelbildungen (statistisch signifikante Kookkurrenzpartner unter den Autosemantika) dient dazu, globale Kontexte und Gemeinsamkeiten sowie Unterschiede im Hinblick auf ihre Distribution in diesen Kontexten zu identifizieren. Im Anschluss daran erfolgt ein Vergleich mit den in den Wörterbüchern eingetragenen Sememen.

- II. Kookkurrenzanalyse II – Disambiguierung von Parallelbildungen im Hinblick auf ihre Bedeutung und Gebrauch (Kookkurrenzanalyse im Bereich ausgewählter attributiver bzw. prädikativer Syntagmen anhand der in Punkt I angeführten Korpusinstrumente, darüber hinaus CCDB – Korpustool zur Kontrastierung naher Synonyme [CNS] von Cyril Belica)

Bedingung – Kookkurrenzanalysen sind nur dann möglich, wenn die zu vergleichenden Formen ausreichend belegt sind. Unter den untersuchten Korpora ist das DeReKo-W am geeignetsten, denn die Trefferanzahl in den DWDS-RZ ist oft zu gering. Für detaillierte Kookkurrenzanalysen sind attributive bzw. prädikative Syntagmen mit denselben syntaktischen Gliedern, die einzelnen Lesarten und globalen Kontexten entsprechen, besonders gut geeignet. Im Rahmen attributiver Syntagmen geht es um vorangestellte adjektivische Attribute, nachgestellte substantivische Genitiv- und Präpositionalattribute (valenzabhängige Komplementen der zugrundeliegenden prädikativen Derivationsbasis), je nach Kontext eignen sich auch koordinative Verknüpfungen von Substantiven mit einem geringeren Fokus, z. B. +3, -3, anschließend werden dann die Frequenzwerte $f_{(a)}$, $f_{(r)}$ und $f_{(p)}$ ermittelt.

Im Rahmen von Detailanalysen hat sich eine Kombination des Substitutionstests, der auf der strukturalistischen und formalsemantischen Tradition der Synonymieforschung beruht (s. dazu Mostýn 2020), mit dem korpuslinguistischen Ansatz, bei dem anhand von Kookkurrenzanalysen mit ausgewählten statistisch signifikanten Kookkurrenzpartnern überprüft wird, ob die Parallelbildung auch tatsächlich gebräuchlich ist, als sehr nutzbringend erwiesen. Im letzten Schritt der Analyse ist noch zu untersuchen, ob areale Unterschiede in der Distribution der Varianten festzustellen ist. Dabei ist die funktionale Auslastung in Texten bundesdeutscher, österreichischer und Schweizer Provenienz ausgehend von den ermittelten Frequenzen zu vergleichen.

Literaturverzeichnis

Primärliteratur

Wörterbücher

Gedruckte Wörterbücher

DUDENREDAKTION (Verf.) (2007): *Duden, Das große Fremdwörterbuch: Herkunft und Bedeutung der Fremdwörter*. 4., aktualisierte Aufl. Mannheim; Leipzig; Wien; Zürich: Dudenverl.

KROME, Sabine (Hrsg.) (2011a): *Brockhaus, Wahrig, Deutsches Wörterbuch*. 9., vollst. neu bearb. und aktualisierte Aufl. Gütersloh; München: Wissenmedia in der Inmedia-ONE-GmbH.

Elektronische Wörterbücher

- ‚DuFWB‘ = Kunkel-Razum, Kathrin (Hrsg.) (2015): *Duden. Das Fremdwörterbuch*. Unentbehrlich für das Verstehen und den Gebrauch fremder Wörter. hrsg. von der Dudenredaktion. 11., vollst. überarb. und aktualisierte Aufl. Berlin: Dudenverl. ISBN 978-3-411-04061-2.
- MATER, Erich (2001): *Rückläufiges Wörterbuch der deutschen Gegenwartssprache*. Vom Autor aktualisierte und wesentl. erw. Ausg. auf CD-ROM. Straelen: Straelener Ms.-Verl.

Online-Wörterbücher

- ‚DOW‘ = Dudenredaktion (o. J.): *Duden Online-Wörterbuch*. Zugänglich unter <https://www.duden.de/> [30. 6. 2023].
- ‚DWDS‘ = *Digitales Wörterbuch der deutschen Sprache*. Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart, hrsg. v. d. Berlin-Brandenburgischen Akademie der Wissenschaften, <<https://www.dwds.de/>> [30. 6. 2023].

Elektronische Korpora und Korpusools

- CCDB: Belica, Cyril (2001–2023): *Kookkurrenzdatenbank CCDB*. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs. © 2001–2023. Institut für Deutsche Sprache, Mannheim. Zugänglich unter: <http://corpora.ids-mannheim.de/ccdb/> [30. 6. 2023].
- DeReKo: Leibniz-Institut für Deutsche Sprache (2022): *Deutsches Referenzkorpus* (DeReKo-2022-1). W-öffentlich – alle öffentlichen Korpora des Archivs W (mit Neuakquisitionen). Zugänglich unter: <https://cosmas2.ids-mannheim.de/cosmas2-web> [30. 6. 2023].
- DWDS-DZ: *Referenz- und Zeitungskorpora des Digitalen Wörterbuchs der deutschen Sprache*. Zugänglich unter <https://www.dwds.de> [30. 6. 2023].
- GBKD: *Google Books Korpus* (Deutsch) (2022). Zugänglich unter: <https://books.google.com/> [30. 6. 2023].
- Google Books NGram Viewer*. Zugänglich unter <https://books.google.com/ngrams> [30. 6. 2023].
- WUL-DN: Leipzig Corpora Collection (2018): *Deutsches Nachrichten-Korpus basierend auf Texten gecrawlt 2018*. Leipzig Corpora Collection. Dataset. Zugänglich unter: https://corpora.uni-leipzig.de/de?corpusId=deu_newscrawl-public_2018. [30. 6. 2023].

Sekundärliteratur

- BELICA, Cyril (1995): *Statistische Kollokationsanalyse und -clustering. Korpuslinguistische Analysemethoden*. © 1995 Institut für Deutsche Sprache, Mannheim. Zugänglich unter <http://corpora.ids-mannheim.de/ccdb/> [30. 6. 2023].
- BELICA, Cyril (2005): *Modellierung semantischer Nähe. Analyse und topografische Visualisierung von Verwendungsaspekten in Self-Organizing-Maps. Korpusanalytische Methode*. Zugänglich unter <http://corpora.ids-mannheim.de/ccdb/> [30. 6. 2023].
- BELICA, Cyril / KEIBEL, Holger / KUPIETZ, Marc / PERKUHN, Rainer / VACHKOVÁ, Marie (2009): Putting Corpora into Perspective. Rethinking synchronicity in corpus linguistics. In: MAHLBERG, Michaela / GONZÁLEZ-DÍAZ, Victorina / SMITH, Catherine (eds.): *Proceedings of the Corpus Linguistics Conference 2009*. Zugänglich unter: <https://ids-pub.bs-zw.de/frontdoor/index/index/docId/4739> [30. 6. 2023].
- BRÜCKNER, Dominik (2009): Die Google-Buchsuche als Hilfsmittel für die Lexikographie. In: *Sprachreport*. Heft 3, Mannheim, S. 26–31.
- BRÜCKNER, Dominik (2012): Noch einmal: Die Google-Buchsuche. In: *Sprachreport*. Heft 2, Mannheim, S. 16–20.
- DIDAKOWSKI, Jörg / GEYKEN, Alexander (2012): *From DWDS Corpora to a German Word Profile – Methodological Problems and Solutions*. Zugänglich unter: https://www.dwds.de/static/publications/pdf/didakowski_geyken_internetlexikografie_2012_final.pdf [30. 6. 2023].
- GEYKEN, Alexander (2011): Statistische Wortprofile zur schnellen Analyse der Syntagmatik in Textkorpora. In: ABEL, Andrea / ZANIN, Renata (Hrsg.): *Korpora in Lehre und Forschung*. Bozen; Bolzano: Universitätsverlag, S. 129–154.
- GOLDHAHN, Dirk / ECKART, Thomas / QUASTHOFF, Uwe (2012): Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages. In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, S. 759–765. Zugänglich unter: http://www.lrec-conf.org/proceedings/lrec2012/pdf/327_Paper.pdf [30. 6. 2023].
- HAHN, Marion (2002): *Die Synonymenlexikografie vom 16. bis zum 20. Jahrhundert. Historische Entwicklung und kommentierte Bio-Bibliografie*. Heidelberg: Winter.
- HERINGER, Hans Jürgen (2012): *Chunking. Synonymik des Deutschen korpusbasiert*. Tübingen: Narr.
- HOLLER, Anke / SCHERER, Carmen (2010): Zur Argumentstruktur entlehnter Verben. In: SCHERER, Carmen / HOLLER, Anke (2010): *Strategien der Iso-*

- lation und Integration nicht-nativer Einheiten und Strukturen*. Berlin; New York: De Gruyter, S. 183–198.
- HÜMMER, Christiane (2007): *Synonymie bei phraseologischen Einheiten*. Frankfurt a. M. u. a.: Lang.
- KEIBEL, Holger / BELICA, Cyril (2007): CCDB: *A Corpus-Linguistic Research and Development Workbench*. Proceedings of the 4th Corpus Linguistics conference. Birmingham. Zugänglich unter: <http://corpora.ids-mannheim.de/cl2007-134.pdf> [30. 6. 2023].
- KRATOCHVÍLOVÁ, Iva (2011): *Kollokationen im Lexikon und im Text: Mehrwortverbindungen im Deutschen und Tschechischen*. Berlin; Münster: Lit.
- MARKOVÁ, Věra (2012): *Synonyme unter dem Mikroskop. Eine korpuslinguistische Studie*. Tübingen: Narr.
- MATER, Erich (1965): *Rückläufiges Wörterbuch der deutschen Gegenwartssprache*. Leipzig: VEB Bibliograph. Institut.
- MICHEL, Jean-Baptiste / SHEN, Yuan Kui / AIDEN, Aviva Presser u. a. (2011): Quantitative analysis of culture using millions of digitized books. In: *Science 2011–14 January*, 331, S. 176–182.
- MOSTÝN, Martin (2017): *Diskrimination oder Diskriminierung von Frauen? Vorüberlegungen zur Analyse von deverbalen Derivaten von Verben auf -ier(en)*. In: *Acta Facultatis Philosophicae Universitatis Ostraviensis. Studia Germanistica*. Nr. 20, Ostrava, S. 51–61.
- MOSTÝN, Martin (2019a): Substantivische Synonyme auf *-ierung* und *-ation* aus sprachwissenschaftlicher und didaktischer Perspektive am Beispiel einer Fehleranalyse studentischer Übersetzungen. In: *Aussiger Beiträge: Germanistische Schriftenreihe aus Forschung und Lehre*. Nr. 13 (1), Aussig, S. 53–79.
- MOSTÝN, Martin (2019b): Parallelbildungen auf *-ation* und *-ierung* aus korpuslinguistischer Perspektive – was zeigt der Signifikanztest? In: VAŇKOVÁ, Lenka (Hrsg.): *Das Fachwort in der Tagespresse*. (= Forum für Fachsprachen-Forschung 159). 1. Aufl. Berlin: Frank & Timme, S. 63–96.
- MOSTÝN, Martin (2020): *Identifikation oder Identifizierung von Katastrophenopfern? Parallelbildungen auf -ation und -ierung in Lexikon und Text aus deutsch-tschechischer Perspektive*. In: DURAND, Marie-Laure / LEFÈVRE, Michel / ÖHL, Peter (Hrsg.): *Tradition und Erneuerung: Sprachen, Sprachvermittlung, Sprachwissenschaft. Akten der 26. Fachtagung der Gesellschaft für Sprache und Sprachen GeSuS e.V. in Montpellier, 5.–7. April 2018*. Hamburg: Verlag Dr. Kovač, S. 121–133.
- MOSTÝN, Martin (2022a): Synonyme in Lexikon und Text: Eine metalexikografische und korpuslinguistische Analyse am Beispiel der Derivate *Stimulation*

- und *Stimulierung*. In: *Acta Facultatis Philosophicae Universitatis Ostravensis. Studia Germanistica*. Nr. 30, Ostrava, S. 5–23.
- MOSTÝN, Martin (2022b): *Deverbale Derivate von Verben auf -ieren – eine meta-lexikografische und korpuslinguistische Analyse*. Habilitationsschrift, unveröffentl. Ostrava.
- MÜLLER, Peter O. (Hrsg.) (2005): *Fremdwortbildung. Theorie und Praxis in Geschichte und Gegenwart*. Frankfurt a. M. u. a.: Lang.
- MÜLLER, Peter O. (2015): Foreign word-formation in German. In: MÜLLER, Peter O. / OHNHEISER, Ingeborg et al. (Hrsg.): *Handbücher zur Sprach- und Kommunikationswissenschaft*. Bd. 40.3. Word-Formation. Berlin; Boston: De Gruyter, S. 1 615–1 637.
- PERKUN, Rainer / BELICA, Cyril (2004): *Eine kurze Einführung in die Konkurrenzanalyse und syntagmatische Muster*. Institut für Deutsche Sprache, Mannheim. Zugänglich unter: <http://www1.ids-mannheim.de/kl/misc/tutorial.html> [30. 6. 2023].
- QUASTHOFF, Uwe / RICHTER, Matthias (2005): Projekt Deutscher Wortschatz. In: *Babylonia* Nr. 3, S. 33–35. Zugänglich unter: http://babylonia.ch/fileadmin/user_upload/documents/2005-3/Baby3_05x.pdf [30. 6. 2023].
- ROGELIO, Nazar / RENAU, Irene (2012): Google Books N-gram Corpus used as a Grammar Checker. In: *Proceedings of the EACL 2012 Workshop on Computational Linguistics and Writing*. Avignon, S. 27–34.
- SEIFFERT, Anja (2002/2005): Probleme synchroner Fremdwortbildungsforschung. In: *Neuphilologische Mitteilungen* 103, S. 161–178. Wiederabgedruckt in: MÜLLER, Peter O. (Hrsg.) (2005): *Fremdwortbildung. Theorie und Praxis in Geschichte und Gegenwart*. Frankfurt a. M. u. a.: Lang, S. 219–239.
- SCHÄUBLIN, Peter (1972): *Probleme des adnominalen Attributs in der deutschen Sprache der Gegenwart*. Berlin; New York: De Gruyter.
- SCHIPPAN, Thea (1967): *Die Verbalsubstantive der deutschen Sprache der Gegenwart*. Habilitationsschrift. Leipzig.
- STORJOHANN, Petra (2006): Kontextuelle Variabilität synonymyer Relationen. In: IDS (Hrsg.): *OPAL – Online publizierte Arbeiten zur Linguistik*. Mannheim. Zugänglich unter: https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/70/file/Storjohann_Kontextuelle_Variabilit%c3%a4t_synonymer_Relationen_2006.pdf [30. 6. 2023].
- STRICKER, Stefanie (2002): Konkurrenzen im Wortbildungssystem um 1800. Aufgezeigt an der Wortbildung Goethes. In: HABERMANN, Mechthild (Hrsg.) (2002): *Historische Wortbildung des Deutschen*. Tübingen: Niemeyer, S. 315–339.
- ŠEMELÍK, Martin (2016): Noch zu *Ge-...(-e)* vs. *-(er)ei*. Eine korpuslinguisti-

sche und lexikographische Betrachtung. In: *Deutsche Sprache* 3/44, Berlin, S. 235–260.

WELLMANN, Hans (1975): *Deutsche Wortbildung. Typen und Tendenzen in der Gegenwartssprache. Zweiter Hauptteil: Das Substantiv*. Düsseldorf: Pädagogischer Verlag Schwann.

WILLEMS, Klaas (2012): Google Books Ngram Viewer und historische Computerlexikologie. In: *Muttersprache 2/2012*, Wiesbaden; Berlin, S. 81–101.

Internetquellen

URL 1: <https://www1.ids-mannheim.de/kl/projekte/korpora/> [30. 6. 2023].

URL 2: <https://www.dwds.de/d/publikationen> [30. 6. 2023].

URL 3: <https://wortschatz.uni-leipzig.de/de> [30. 6. 2023].

URL 4: <https://blog.google/products/search/google-books-library-project/#logistics> [30. 6. 2023].

URL 5: <https://www.dwds.de/b/dwds-wortprofil-in-neuer-version/> [30. 6. 2023].