

Ulrich Schnörch

## Kookkurrenzanalyse und Vergleich: Überlegungen zur Methoden-anwendung bei der lexikografischen Beschreibung von Paronymen

### Abstract

Im Rahmen der korpusgestützten Lexikografie stellt die Kookkurrenzanalyse ein bewährtes Verfahren dar, um Massendaten aus Korpora im Corpus-driven-Ansatz zu einem Einzelstichwort vorzustrukturieren. Wie diese Daten im redaktionellen lexikografischen Prozess in die Wortartikelproduktion einfließen können, wurde beispielsweise beim allgemeinen einsprachigen Online-Wörterbuch *elexiko* erprobt, dokumentiert und umgesetzt. Für das Wörterbuch „Paronyme – Dynamisch im Kontrast“ bildet die Kookkurrenzanalyse gleichfalls einen Ausgangspunkt für die Arbeit an Wortartikeln, allerdings unter anderen Voraussetzungen: Der folgenreichste Unterschied in methodischer Hinsicht ist, dass im Paronymwörterbuch mindestens zwei Stichwörter vergleichend gegenübergestellt werden, um so semantische Ähnlichkeiten und Unterschiede explizit zu machen. Im Beitrag wird diskutiert, wie die Verfahren der Kookkurrenzanalyse und des nachfolgenden Vergleichs für die praktische Artikelarbeit beim Paronymwörterbuch variiert, spezifiziert und nutzbar gemacht wurden.

In the framework of a corpus-driven lexicography, it is the computer-based analysis of collocations which has established itself as a valid method for pre-structuring masses of corpus data on individual lexical units. The general monolingual online dictionary *elexiko* has successfully shown and documented how these data can be used for the compilation of entries in the editorial lexicographic process. For the new dictionary “Paronyme – Dynamisch im Kontrast” (‘Paronyms – Dynamic in Contrast’), co-occurrence analyses also form the basis for lexical descriptions, albeit under different conditions: the most consequential difference in terms of methodology is that the dictionary of paronyms compares at least two lexemes side by side. By doing so, semantic similarities and differences can be made explicit. In this paper, we will discuss how the procedures of co-occurrence analysis and subsequent lexical comparison can vary, and how they are adjusted and employed in the practical process of dictionary-writing.

### 1. Die Kookkurrenzanalyse als korpuslinguistische Methode zur Analyse von Einzelstichwörtern – eine Skizze

Die Kookkurrenzanalyse ist im Rahmen der korpusgestützten Lexikografie ein erprobtes, variables, anpassungsfähiges und vielfältig einsetzbares Verfahren. Es dient grundsätzlich der Vorstrukturierung sprachlicher Massendaten (vgl. Perkuhn/Keibel/Kupietz 2012; Belica 1995). Am Beispiel des einsprachigen, allgemeinen Online-Wörterbuchs *elexiko*<sup>1</sup> soll im ersten Abschnitt exemplarisch in Erinnerung gerufen werden, auf welche Weise die Lexikograf/innen bei der Arbeit an Einzelwortartikeln die Kookkurrenzanalyse nutzen, sprich: auf welche Weise welche Wörterbuchdaten korpusgestützt ermittelt bzw. abgeleitet wurden (vgl. ausführlich dazu Schnörch 2005, 2015). Vor diesem Hintergrund kann im zweiten Abschnitt leichter nachvollzogen werden, auf welche Weise das Verfahren für das mindestens zwei Stichwörter vergleichende Paronymwörterbuch erprobt, abgewandelt und spezifiziert wurde (vgl. auch Perkuhn in diesem Heft).

<sup>1</sup> *elexiko* war das erste deutschsprachige korpusgestützte elektronische Wörterbuch, vgl. [www.elexiko.de](http://www.elexiko.de) (Stand: 6. 11. 2017).

Das Stichwort *Technologie* ist mit 178.748 Treffern (Stand: 14.11.2017) im *lexiko*-Korpus belegt. Selbstredend kann kein Mensch diese Masse sprachlicher Daten völlig unstrukturiert lexikografisch in irgendeiner sinnvollen Form nutzbar machen. Das Verfahren der Kookkurrenzanalyse bietet hingegen eine Möglichkeit, diese Aufgabe einer statistischen Vorstrukturierung (in Form von Auswertungen der kontextuellen Nachbarschaft) automatisch durchführen zu lassen. Das Ergebnis ist eine Liste, die u. a. Kookkurrenzpartner und (dazugehörige) syntagmatische Muster enthält, die sich beispielsweise nach absteigender Signifikanz sortieren lassen. Aus einer solchen Liste lässt sich eine ganze Reihe von Daten direkt für die Arbeit am Wortartikel nutzbar machen: Sie dienen in einem wechselseitigen Prozess gleichzeitig der Stichwortanalyse und der Generierung konkreter Wörterbuchinformationen. Eine kleine Auswahl soll dies illustrieren. Ich beschränke mich hier auf Angaben bzw. Angabetypen, die es – allerdings in abgewandelter Form – auch im Paronymwörterbuch gibt (siehe unten).

Als erstes lassen sich die **Kollokatoren** selbst auswählen und nach bestimmten, beispielsweise semantischen Gesichtspunkten sortieren. Für das Beispiel *Technologie* können etwa Gruppen gebildet werden wie

- *alternativ, ausgereift, effizient, energiesparend* usw., *anwenden, beherrschen, einsetzen, entwickeln*. Diese und zahlreiche andere Adjektive und Verben weisen auf eine Einzelbedeutung im Sinne von ‚Produktionstechnik‘ hin.
- *modern, veraltet* usw., *bauen, besitzen, exportieren, kaufen*. Diese und zahlreiche andere Adjektive und Verben passen eher zu einer Einzelbedeutung im Sinne von ‚technisches Gerät‘.

Es ist möglich, **syntagmatische Muster** auszuwählen, die als signifikante, wortübergreifende Einheiten gewissermaßen als Satzbausteine fungieren können, z. B.

- *auf saubere Technologien setzen, mit westlicher Technologie* u. a., die wieder der Einzelbedeutung ‚Produktionstechnik‘ zugeordnet werden können.
- *die neueste Technologie auf den Markt bringen, die Technologie an China verkaufen, mit der Technologie vertraut sein* u. a., die zur Einzelbedeutung ‚technisches Gerät‘ passen.

Ferner finden sich unter den Kollokatoren auch Ausdrücke, die sinnrelational mit dem Stichwort in Verbindung stehen, etwa in **Beziehungen der Ähnlichkeit**, z. B. *Produktionsmethode, Verfahren, Technik* u. a. synonyme Ausdrücke zu *Technologie* im Sinne von ‚Produktionstechnik‘ sowie *Gerät, Maschine* als Synonyme zu *Technologie* im Sinne von ‚technisches Gerät‘.

Der erste Schritt einer automatischen Vorstrukturierung bildet also die Grundlage für einen sehr wesentlichen zweiten Arbeitsprozess, der vorrangig redaktionell durchzuführende (interpretatorische) Analysen umfasst. Hierbei muss bedacht werden, dass jede Auswahl und Sortierung – zumal mit lexikografischen Zielsetzungen – gewisser Vorgaben bedarf, wie sie beispielsweise in einem Redaktionshandbuch erarbeitet und zusammengefasst werden. Dabei handelt es sich um methodische Richtlinien, welche etwa die Auswahl und Sortierung steuern und damit transparent machen.

Die zweite Gruppe an Informationen, die der Kookkurrenzliste entnommen wird, ist nur insofern von der ersten abzuheben, als bei ihr redaktionell-analytische Prozesse stärker in den Vordergrund rücken. Hierzu gehört beispielsweise die semantische Analyse, welche

freilich – wie oben bereits angedeutet – die skizzierten Auswahl- und Sortierungsvorgänge beeinflusst: Für die lexikografische Beschreibung ist es unabdingbar, die ‚Bedeutungen‘ der Wörter zu analysieren, egal, ob man sie **Lesarten** (wie in *elexiko*) oder **Verwendungsaspekte** (wie im Paronymwörterbuch) nennt. Insofern ist bereits die Gruppierung der Kollokatoren ein erster Schritt in diese Richtung: Aus semantisch ‚kohärenten‘ Gruppen von Kollokatoren lassen sich Einzelbedeutungen ableiten: Eine Gruppe entspricht einer Einzelbedeutung. Diese korpusgestützte Ermittlung von Lesarten wurde in *elexiko* ergänzt durch die Zuweisung einer Bedeutung in die Kategorie einer Prädikatorenklasse (vgl. Haß 2005). Die beiden oben angeführten Lesarten von *Technologie* wären Individuativa; daneben gibt es Handlungs-, Eigenschafts- und weitere Prädikatorenklassen. Für *Technologie* wurden mit diesem Verfahren insgesamt fünf Lesarten in *elexiko* ermittelt. Erst die auf dem Wege dieses Wechselspiels von Auswahl, Gruppierung und Kategorisierung vorgenommene semantische Differenzierung ermöglicht(e) es in weiteren überwiegend redaktionell ablaufenden Arbeitsschritten, die Kollokatoren innerhalb der fünf Lesarten in einem weiteren Schritt nach satzsemantischen Gesichtspunkten zu subkategorisieren und aus der Kookkurrenzliste gewonnene syntagmatische Muster und sinnrelationale Ausdrücke lesartenbezogen zuzuordnen.

In Einzelfällen ist es auch möglich, aus den Kookkurrenzlisten Indizien für **themen- oder fachsprachenspezifischen Gebrauch** abzuleiten und ggf. entsprechende **Gebrauchsbesonderheiten**, z. B. als themen- oder diskursspezifischen Gebrauch bzw. ‚Domänenangabe‘, lexikografisch zu dokumentieren. Bei *Technologie* war das in *elexiko* allerdings nicht der Fall.

Die redaktionelle Auswertung von Kookkurrenzlisten war der zentrale Vorgang bei der Artikelarbeit in *elexiko* mit dem Ziel, ein einzelnes Stichwort möglichst facettenreich und umfassend zu beschreiben. Im neuen Paronymwörterbuch konnte bei der lexikografischen Arbeit auf eine Reihe von dort gemachten Erfahrungen zurückgegriffen werden. Folgende Punkte seien nochmals zusammengefasst, da sie Eigenheiten von *elexiko* darstellen, die beim Paronymwörterbuch anders gehandhabt werden:

- Ein einzelnes Stichwort ist der Ausgangspunkt.
- Die semantische Analyse (Lesartendisambiguierung) erfolgt mit dem Ziel, Einzelbedeutungen zu ermitteln, deren Kategorisierung auf referenziellen Unterschieden und der Kategorisierung in Prädikatorenklassen beruht.
- Die lexikografische Beschreibung der einzelnen Lesarten bezüglich ihrer sprachlichen Verwendungsbedingungen erfolgt möglichst facettenreich und detailliert (was beispielsweise die Angabe zahlreicher Kollokatoren und sinnrelationaler Ausdrücke belegt). Sie bleibt i. d. R. auf sprachliche Informationen beschränkt.

Wollen Nutzer/innen sich über den Gebrauch von *Technologie* informieren, finden sie in *elexiko* auf der entsprechenden Artikelseite fünf Lesarten als Einstieg für weitere Informationen.

Der abgebildete Ausschnitt der Einstiegsseite des *elexiko*-Artikels *Technologie* mag die einführenden Überlegungen zur Kookkurrenzanalyse und zu lexikografischen Arbeitsprozessen beenden. Vor diesem Hintergrund ist es nun einfacher, einige Besonderheiten des Paronymwörterbuchs vergleichend herauszuarbeiten sowie Besonderheiten seiner Konzeption und methodischen Herangehensweise zu beschreiben.

elexiko

## Technologie 🔊

---

Lesartenübergreifende Angaben

**④ Orthografie**  
 Normgerechte Schreibung: Technologie  
 Worttrennung: Tech|no|logie

**④ Wortbildung**  
 Wortbildungsart/-typ: explizite Derivation  
 Basis: Technik (Nomen)  
 Suffix: -(o)logie

**④ Wortbildungsprodukte**  
 (automatisch ermittelt) weiter »

---

Lesartenbezogene Angaben ⓘ

Lesart **'Produktionstechnik'** weiter »  
 Mit *Technologie* bezeichnet man eine Technik und die damit verbundenen Verfahren und Methoden, die zur Herstellung oder Bearbeitung eines Produktes oder für dessen Funktionstüchtigkeit nötig sind.

Lesart **'technisches Gerät'** weiter »  
 Mit *Technologie* bezeichnet man ein technisches, meist elektronisches Gerät (z. B. ein Handy) oder Arbeitsmittel (z. B. Computer).

Lesart **'Methode'** weiter »  
 Mit *Technologie* bezeichnet man eine Methode oder eine Praktik des Vorgehens bzw. eine Fertigkeit, die zur Ausübung von etwas notwendig ist.

Lesart **'Bereich'** weiter »  
 Mit *Technologie* bezeichnet man einen Bereich, der sich mit Fragen der Produktionstechnik beschäftigt.

Abb. 1: Das Stichwort *Technologie* in *elexiko*

## 2. Die Kookkurrenzanalyse zur vergleichenden Untersuchung von Paronymen

Sucht man im Paronymwörterbuch nach *Technologie*, so deutet bereits der erste optische Eindruck (Abb. 2) an, dass dort ein völlig anderer Ansatz umgesetzt wird.<sup>2</sup> Offensichtlich steht nicht ein isoliertes Stichwort und dessen möglichst umfassende Beschreibung im Zentrum des lexikografischen Interesses, sondern der direkte Vergleich von mindestens zwei Paronymen. Der vollständige Titel lautet nicht von ungefähr „Paronyme – Dynamisch im Kontrast“. Er ist zugleich Programm: Im Folgenden soll veranschaulicht werden, inwiefern die beiden lexikografischen Anforderungen Wörter zu kontrastieren und dynamisch zu präsentieren, die redaktionelle Auswertung der Kookkurrenzdaten beeinflussen.

Auch das Paronymwörterbuch erarbeitet(e) seine Daten/Informationen von der Erstellung der Stichwortliste bis hin zu den Informationen in den Wortartikeln auf Korpusbasis. Die Lexikograf/innen bedienen sich bei der Artikelarbeit der Kookkurrenzanalyse, im Paronymwörterbuch allerdings einem von mehreren sich im Idealfall ergänzenden korpusanalytischen Verfahren (vgl. dazu auch Teichmann in diesem Heft). Im Zuge der Kookkurrenzanalyse führt der Weg über redaktionelle lexikografische Analysen, Auswahl-, Gruppierungs- und Strukturierungsprozeduren hin zur Ableitung abstrakterer Informatio-

<sup>2</sup> Der Screenshot vermittelt nur ein rein statisches Bild des Paronymwörterbuchs. Im Gegensatz zu *elexiko* spielen dynamische und daher variable Präsentationsformen der Daten eine zentrale Rolle (vgl. Storjohann im vorliegenden Heft).

nen.<sup>3</sup> Die im ersten Abschnitt vorgestellten Angabetypen (Kollokatoren, syntagmatische Muster, sinnrelationale Ausdrücke) kommen auch im Paronymwörterbuch vor, allerdings in z. T. stark abgewandelter Form: Rein quantitativ erfolgt eine deutlich selektivere Auswahl. Qualitativ liegen den namentlich gleichen Angaben bzw. Angabetypen andere Konzepte zugrunde.

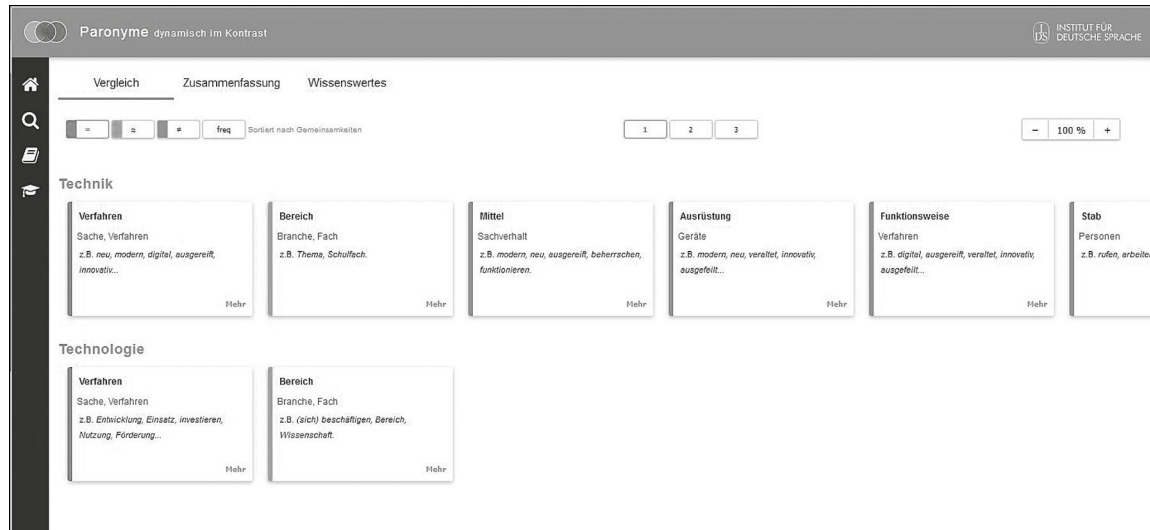


Abb. 2: Das Stichwort *Technik/Technologie* im Paronymwörterbuch

Struktureller Dreh- und Angelpunkt des Paronymwörterbuchs ist also der Vergleich: Paronyme lexikografisch zu beschreiben, bedeutet, zwei oder mehr Stichwörter in Kontrast zu setzen. Vergleiche begegnen einem ständig, auch im Alltag. Im Volksmund heißt es lapidar, man könne oder solle keine Äpfel mit Birnen vergleichen. Das scheint auf den ersten Blick sehr plausibel, auf den zweiten bringt es uns zu Grundvoraussetzungen des methodischen Vergleichens:

Das Vergleichsverfahren ist nicht nur nützlich, sondern unabdingbar. Und ob zwei zu vergleichende Objekte identisch oder gänzlich verschieden sind, läßt sich erst nach einem Vergleich feststellen. [...] Jeder Vergleich wird erst durch ein 'tertium comparationis' sinnvoll. Ein 'tertium comparationis' legitimiert auch scheinbar sinnlose Vergleiche. Landläufige Redensarten wie 'Man kann Äpfel nicht mit Birnen vergleichen' sind so nicht haltbar. Man kann Äpfel sehr wohl mit Birnen vergleichen, z. B. im Hinblick auf die 'tertia comparationis' Lagerfähigkeit, Marktpreis oder – wie häufig in ernährungswissenschaftlichen Tabellen – auf den Vitamingehalt. (Spillner 1997, S. 209f.)

Es ist wichtig, sich bewusst zu machen, dass für einen Vergleich ganz allgemein die Vergleichsbedingungen elementar sind. Konkret auf das Paronymwörterbuch übertragen bedeutet dies, sinnvolle Vergleichsparameter zu finden und methodische Transparenz zu schaffen. Für die lexikografische Beschreibung relational verbundener Stichwörter, also von mindestens zwei Paronymen, ist ein Vergleich inhaltlich wie formal maßgebend. Aufgabe des Wörterbuchs ist es, diese Vergleichbarkeit nachvollziehbar her- und darzustellen. Dazu müssen die Daten der Wortartikel entsprechend ausgewählt, redaktionell bearbeitet

<sup>3</sup> Dabei kondensieren und kodieren Lexikograf/innen schrittweise nach bestimmten redaktionellen Vorgaben Daten zu strukturierten Informationen unterschiedlicher Abstraktionsebenen. Diese granularen Informationen müssen Nutzer/innen in umgekehrter Richtung dekodieren, um angemessene Einordnungen bzw. Bezüge herstellen, also sinnvolle Interpretationen vornehmen zu können.

und schließlich präsentiert werden. Veranschaulichen wir das Gesagte nun anhand der Paronyme *Technik/Technologie* bzw. *technisch/technologisch*.

Die Recherche im Paronymkorpus und die Kookkurrenzanalyse zu den betreffenden Stichwörtern führen zu folgenden Ergebnissen:

*Technik*: 199.422 Treffer im Korpus, 3.574 Kookkurrenzen

*Technologie*: 48.369 Treffer im Korpus, 2.839 Kookkurrenzen

*technisch*: 282.122 Treffer im Korpus, 3.327 Kookkurrenzen

*technologisch*: 12.518 Treffer im Korpus, 1.112 Kookkurrenzen

Es ist wichtig, solche Zahlen stets im Hinterkopf zu behalten, weil sie im fortschreitenden Verlauf der Analyse möglicherweise Interpretationsansätze für diesen oder jenen Befund liefern können. Bei unseren Beispielen differieren sie zum Teil erheblich. Den weiteren Vergleich beeinträchtigt dieser Umstand jedoch nicht wesentlich. Die Kookkurrenzlisten beinhalten trotzdem noch ausreichend signifikante Kollokatoren, um die semantische Analyse der Paronyme zu ermöglichen und die Auswahl und Ableitung weiterer Wörterbuchdaten zu gewährleisten. Es gibt allerdings Beispiele, bei denen die wenigen Treffer eines Paronyms die Aussagekraft einer Kookkurrenzanalyse stark einschränkt oder diese im Extremfall gänzlich unmöglich macht.<sup>4</sup> Die Vergleichsbasis gerät dann rein quantitativ in Schiefelage, was allerdings durch die Formulierung des entsprechenden Befundes im Wörterbuch, also der Modifizierung der Vergleichsbedingungen, etwas aufgefangen und für Nutzer/innen deutlich gemacht werden kann.

Auswahl und Gruppierung der **Kollokatoren**: Sie erfolgt mit dem wesentlichen Unterschied einer deutlichen quantitativen Beschränkung. Es werden für die Präsentation<sup>5</sup> maximal und exemplarisch die zehn signifikantesten ausgewählt und in absteigender Signifikanz aufgelistet. Es besteht die Möglichkeit mehrere solcher Gruppen zu erstellen, deren Mitglieder dann im weitesten Sinne aufgrund ihrer satzsemantischen Funktion zusammengehören, also etwa Handlungen, Eigenschaften u. a., z. B.:

**Technik** ‚Produktionsverfahren, -weise‘

- *neu, modern, digital, ausgereift, innovativ, zukunftsweisend*
- *erlernen, beherrschen, entwickeln, funktionieren*

**Technologie** ‚Produktionsverfahren, -weise‘

- *umweltfreundlich, innovativ, zukunftsweisend, nachhaltig, veraltet, ausgereift, aufwendig, biometrisch*
- *einsetzen, ersetzen, fördern*

Auswahl und Gruppierung der **syntagmatischen Muster**: Sie erfolgt gleichfalls in quantitativ deutlich reduzierter Form, die Obergrenze sind fünf, wobei die Möglichkeit besteht, Varianten anzugeben. Das signifikanteste Muster steht an erster Stelle, gefolgt von den zunehmend weniger signifikanten. Zur Illustration seien hierfür die Kontextmuster von *Technologie* ‚Produktionsverfahren/-weise‘ angeführt:

<sup>4</sup> Etwa Fälle wie *antikisch*: 157 (!) Treffer / *antik*: 36.328, *schuldfrei*: 42 (!) Treffer / *schuldlos*: 1.683 (siehe dazu auch Reckenthäler in diesem Heft).

<sup>5</sup> Die quantitative Beschränkung betrifft natürlich nur die lexikografisch-dokumentarische Seite. Zur linguistisch-lexikografischen Analyse des Stichwortes wird selbstverständlich die gesamte Kookkurrenzliste herangezogen.

- in [umweltfreundliche/zukunftsträchtige] Technologien investieren,
- Technologien für den intelligenten Verkehr,
- Technologien der Zukunft,
- die Technologie ist noch nicht ausgereift,
- Technologien und [Produkte/Dienstleistungen/Märkte].

Auswahl (und Gruppierung) **sinnrelationaler Ausdrücke**: Sie beschränkt sich auf die Angabe von Synonymen, manchmal auch Antonymen, wobei innerhalb eines Verwendungsaspektes jeweils nur maximal fünf der signifikantesten mit entsprechenden Korpusbelegen dokumentiert werden.

Auswahl von **Korpusbelegen**: Maximal drei dienen zur Illustration eines Verwendungsaspektes, jeweils einer veranschaulicht die Relation zwischen dem paronymen Stichwort und einem Synonym bzw. Antonym (siehe oben).

Die Ermittlung dieser lexikografischen Daten zu einem Paronym ist einerseits Verfahren und Ergebnis des Analysevorgangs. Andererseits dient die Gegenüberstellung dieser Daten mit denen des zweiten paronymen Stichworts gleichzeitig als erste Stufe des Vergleichs. Was auf den ersten Blick als lineare Folge lexikografischer Arbeitsschritte anmutet, vollzieht sich tatsächlich unter vielfachen Rückkopplungsprozessen und endet schließlich mit dem ‚fertigen‘ Wörterbucheintrag. Die Kookkurrenzanalyse ist der Grundstein für die wechselseitig organisierten Prozesse der Einzelparonymanalyse und des Paronymvergleichs; erstere (und damit auch letztere) wird schrittweise von der rein sprachlichen Ebene bis hinauf zur kognitiven fortgesetzt, z. B. in der Formulierung von sog. „Guide Words“ (siehe unten). Je nachdem, ob und mit welchem Deckungsgrad sich Schnittmengen der Datengruppen bilden lassen, werden drei Gruppen von Relationen zwischen den Einzelbedeutungen unterschieden: eine mit ‚starken semantischen Überlappungen‘, eine ‚semantisch ähnliche‘ und eine, die ‚individuelle Verwendungsaspekte ohne Ähnlichkeit‘ beinhaltet. Die radikale Auswahl aus der Materialfülle (z. B. ein Pool von 3.574 Kollokatoren bei *Technik*) ist für die Lexikograf/innen im Einzelfall nicht immer problemlos. Das Kriterium der Signifikanz/Frequenz bietet jedoch i. d. R. eine verbindliche Entscheidungshilfe nach dem Motto „was zuerst kommt, wird auch dokumentiert“.<sup>6</sup>

Die quantitative Beschränkung der Artikeldaten ist die Voraussetzung, um den lexikografischen Vergleich erst zu ermöglichen und diese Gegenüberstellung auch auf möglichst einen Blick überschaubar zu machen, also Gemeinsamkeiten und Unterschiede akzentuiert, d. h. datenfundierte Präsentationen zu können. Bei der Konzeption des Paronymwörterbuchs wurde von Anfang an größter Wert darauf gelegt, dass Nutzer/innen (nach einer gewissen Einstiegsphase) die stark kondensierten und kodierten Wörterbuchttexte schnell und intuitiv erfassen, dekodieren können.<sup>7</sup> Vor dem Hintergrund des Vergleichs bedeutet

<sup>6</sup> Ausnahmen von dieser Regel sind denkbar, wenn beispielsweise mehrheitlich Komposita mit gleichem Grundwort die ersten Kollokatorenreihen belegen. Um das Gebrauchsspektrum in seiner Vielfalt zu illustrieren, mag es dann sinnvoll erscheinen, ähnlich gebildete Komposita durch andersartige, nachrangige Kollokatoren zu ersetzen. Gleichzeitig wird damit das Kriterium der Auswahl nach statistischem Rang durch das der Auswahl nach lexikografischer Relevanz ersetzt.

<sup>7</sup> Im *lexiko*-Artikel *Technologie* wurden seinerzeit zur möglichst umfassenden Beschreibung dieses einen Stichwortes fünf Lesarten unterschieden und diesen insgesamt knapp 120 Kollokatoren zugeordnet, die sich wiederum in 16 Gruppen einteilen ließen. Es wurden insgesamt über 30 Konstruktionen angegeben und ca. 50 sinnrelationale Ausdrücke jeweils mit Beleg illustriert. Ein Gedankenspiel: Ein expliziter

das in Bezug auf den Datenumfang die Umsetzung der Vorgabe „weniger ist mehr“. Die Auswahl und Generierung der konkreten Wörterbuchdaten geht selbstverständlich auch im Paronymwörterbuch mit der korpusgestützten, semantisch-diskursiv ausgerichteten redaktionellen Analyse einher. Es sind – wie mehrfach schon betont – zwei Seiten einer Medaille; das eine, die Analyse, bedingt das andere, die Dokumentation der Ergebnisse, und umgekehrt. Erst dieser fortlaufende Rückkopplungsprozess garantiert, dass selbst bei starker Auswahl lexikografischer Daten die Analysebasis ausreichend breit bleibt.

Die redaktionelle Auswertung der Kookkurrenzanalyse, also die Sortierung der Kollokatoren, bietet auch die Möglichkeit, unterscheidbare Verwendungskontexte eines Lemmas herauszumodellieren. Zur (semantischen) Differenzierung der Stichwörter im Paronymwörterbuch schreibt Storjohann (in diesem Heft in Abschn. 3.1, S. 86):

Diese Verwendungskontexte entsprechen nicht zwangsläufig traditionellen Lesarten, sondern sollen bestimmte Aspekte kommunikativ-situativen Gebrauchs demonstrieren. Daher wird auf herkömmliche Disambiguierungsprinzipien verzichtet, vielmehr rechtfertigen divergierende Referenz bzw. konzeptuelle Kategorien eine Trennung. [...] In kontrastiv angelegten Kategorien werden die Arten der Verwendungsaspekte eines jeden Paronyms zu denen des Partnerausdrucks in Beziehung gesetzt.

Die Verwendungsaspekte bilden also die lexikografische (und semantisch-kommunikative) Bezugseinheit, die mithilfe der Kookkurrenzliste ermittelt werden kann, genauer: mittels der Gruppierung von Kollokatoren. Diese Gruppierung erfolgt allerdings nicht ausschließlich mit Bezug auf deren semantische Nähe, sondern zusätzlich auf einer etwas anderen, über die Dimension des rein Sprachlichen hinausreichenden Ebene, nämlich der ihrer kommunikativen Situierung. Zur Ermittlung von Verwendungsaspekten können neben dieser Möglichkeit auch andere korpusanalytische Verfahren nutzbar gemacht werden (vgl. Teichmann in diesem Heft). Jede Kurzparaphrase wird durch ein Bezugswort, ein sog. Guide Word, ergänzt, „einer Kategorienbezeichnung, die den referenziellen Kontext in einem Bezugsrahmen einbettet“ (Storjohann in diesem Heft in Abschn. 3.1, S. 86). Die Ableitung dieser Guide Words erfolgt auf dem Weg der Abstrahierung aus einer signifikanten Kollokatorengruppe des entsprechenden Verwendungsaspektes. Bei Adjektiven geht dies verhältnismäßig problemlos. Bei *technisch* ‚die Arbeitsverfahren betreffend‘ sind beispielsweise folgende Kollokatoren zu einer Gruppe zusammengefasst, die den nominalen Bezug zum Adjektiv im Korpus konkretisiert bzw. illustriert: *Fortschritt, Innovationen, Entwicklungen, Kompetenz, Lösungen, Überlegenheit, Veränderungen, Herausforderungen, Trend, Machbarkeit*; aus dieser Kollokatorengruppe lässt sich leicht das Guide Word SACHVERHALT abstrahieren.

Ein weiterer Verwendungsaspekt ist z. B.:

‚Einrichtung betreffend‘:

SACHVERHALT: *Defekt, Fehler, Probleme, Mängel, Berufe, Details, Aufwand, Schwierigkeiten, Perfektion, Panne, Hilfe.*

AUSRÜSTUNG: *Ausstattung, Geräte, Anlagen, Ausrüstung, Hilfsmittel, Equipment, Gebäudeausrüstung.*

---

Vergleich dieses *ellexiko*-Wortartikels mit einem vergleichbaren Wortartikel *Technik* (dessen redaktionell bearbeitete Fassung in *ellexiko* allerdings nicht existiert) ist lexikografisch kaum umsetzbar und vor allem auf dem Bildschirm eines Tablett oder Smartphones nicht realisierbar.

Die redaktionelle Auswertung der Kookkurrenzliste führte im Falle von *Technologie* zur Unterscheidung von vier Verwendungsaspekten, im Falle von *Technik* zu sieben Kontexten.

Bei Nomina gestaltet sich die Formulierung eines kognitiv-konzeptionell aufzufassenden Bezugswortes etwas komplizierter, da auf syntagmatischer Ebene allenfalls indirekte Bezüge hergestellt werden. Für *Technik* ‚Produktionsverfahren, -weise‘ gibt es zwei Kollokatorengruppen, nämlich:

„Wie wird Technik charakterisiert?“

*neu, modern, digital, ausgereift, innovativ, zukunftsweisend;*

„Was macht man in Bezug auf Technik?“

*erlernen, beherrschen, entwickeln, funktionieren.*

Die mehr oder minder direkte Erschließung eines Guide Words ist in diesem Fall kaum möglich. Um dennoch mittels eines operationalisierbaren Verfahrens ein Bezugswort abzuleiten, wird momentan mit Hilfsformulierungen experimentiert, also auf *Technologie* ‚Produktionsverfahren‘ angewandt: *mit Bezug auf/im Sinne von* HERSTELLUNG (= Guide Word). Weitere Angaben, die für ein (einzelnes) Paronym korpusgestützt, d. h. maßgebend auf Basis der Kookkurrenzanalyse mehr oder minder indirekt abgeleitet werden, sind u. a. die Langparaphrase und ggf. die Angabe eines diskursiv dominanten Verwendungsbereichs. Die Quervernetzung der Daten innerhalb des Wortartikels ist durch die Rückkopplung an das Korpus gewährleistet. Sie lässt sich besonders gut am Beispiel der Paraphrasenbelege zeigen. Diese werden bewusst so ausgewählt, dass sie i. d. R. einen Kollokator aus der Kollokatorengruppe enthalten. Die Wortartikel des Paronymwörterbuchs enthalten folgende Daten, die online jedoch völlig anders präsentiert werden:

<b>1. Technik</b>		<b>2. Technologie</b>
Produktionsverfahren, -weise	<b>Kurzparaphrase</b>	Produktionsverfahren, -weise
Herstellung	<b>Bezugswort</b>	Herstellung
bezeichnet eine Methode bzw. ein planvoll durchgeführtes Verfahren zur (industriellen) Herstellung oder Bearbeitung eines Produktes bzw. zur Bereitstellung einer Dienstleistung	<b>Langparaphrase</b>	bezeichnet eine Methode bzw. ein planvoll durchgeführtes Verfahren zur (industriellen) Herstellung oder Bearbeitung eines Produktes bzw. zur Bereitstellung einer Dienstleistung
meist in <b>Wirtschaft</b>	<b>diskursiver Verwendungsbereich</b>	meist in <b>Wirtschaft</b>
Wie wird Technik charakterisiert? <i>neu, modern, digital, ausgereift, innovativ, zukunftsweisend.</i> Was macht man in Bezug auf eine Technik? <i>erlernen, beherrschen, entwickeln, funktionieren.</i>	<b>Kollokationen</b> [nochmals subkategorisiert nach satzsemantischen Funktionen]	Wie wird Technologie charakterisiert? <i>umweltfreundlich, innovativ, zukunftsweisend, nachhaltig, veraltet, ausgereift, aufwendig, biometrisch.</i> Was macht man mit Technologie? <i>einsetzen, ersetzen, fördern.</i> Welche Tätigkeiten werden in Bezug auf Technologie assoziiert? <i>Einsatz, Entwicklung, Nutzung, Förderung, Einführung.</i>

1. <i>Technik</i>		2. <i>Technologie</i>
<i>Methode, Technologie, Verfahren</i>	<b>Synonyme</b> [jeweils mit Beispielen belegt]	<i>Methode, Technik, Verfahren</i>
Was in Michelbach mit modernster <b>Technik</b> und höchstem Qualitätsstandard produziert wird, das vollführen die Kinder auf ganz handwerkliche Art zur Zeit im Kindergarten. So wird Getreide zu Flocken gequetscht und Mehl von Hand gemahlen. (Rhein-Zeitung, 20.10.2005, Kinder erlebten, wie das Mehl entsteht.)	<b>Belege</b> [hier nur der erste ausgewählt, U. Sch.]	Der neue Generalsekretär der Deutschen Bundesstiftung Umwelt (DBU), Heinrich Bottermann, hat sich gegen die umstrittene Erdgasfördermethode Fracking ausgesprochen. „Die <b>Technologie</b> ist fragwürdig“, sagte er am Mittwoch. (die tageszeitung, 10.10.2013, S. 22, Neuer Chef lehnt Fracking ab.)

Tab. 1: Für *Technik/Technologie* im Paronymwörterbuch aus den Kookkurrenzlisten ermittelte Daten

Die wenig elegant anmutende Präsentation dieser Daten in Tabelle 1 wurde bewusst gewählt: Sie visualisiert den Vergleich zweier semantisch ähnlicher Verwendungsaspekte von *Technologie* und *Technik* zwar nicht in optimaler Weise. Dafür spiegelt sie das Analyseverfahren mittels Kookkurrenzanalyse bzw. dessen Ergebnis insofern wider, als diese stets Stichwort (und dessen Verwendungsaspekt(e)) für Stichwort (und dessen Verwendungsaspekt(e)) nacheinander erfolgt, d. h. es ist stets nur möglich, jedes Lemma isoliert zu untersuchen. Die eigentliche Vergleichsprozedur erfolgt durch die Lexikograf/innen dann auf redaktionellem Weg. Damit ist der Hauptnachteil des Kookkurrenzverfahrens bei der Anwendung als lexikografische Methode für Paronyme formuliert: Sie unterstützt keine (paronym-) vergleichenden Automatismen bei der Strukturierung sprachlicher Korpusdaten. Der im Wortartikel angestrebte Vergleich von Paronymen ist also stets das Ergebnis redaktioneller Arbeit der Lexikograf/innen, die ihren Ausgangspunkt in zwei (oder mehr) Kookkurrenzlisten nimmt.

Für das Beispiel *Technologie* und *Technik* wurden neben den beiden bereits angeführten jeweils ein weiterer Verwendungsaspekt mit starker semantischer Ähnlichkeit (,Produktionsverfahren‘) ermittelt, daneben aber auch jeweils weitere, die kein semantisches Pendant beim Partnerwort besitzen, z. B.

**Technik** ‚Personengruppe‘

[Bezugswort] Fachexperte(n)

bezeichnet eine (organisierte) Gruppe von Personen, die für den Einsatz, die Bedienung und Steuerung von Geräten verantwortlich ist.

Kollokationen

Was macht die Technik?

*arbeiten, sorgen, anrücken.*

Der (Daten-)Vergleich mit *Technologie* zeigt dort eine ‚Leerstelle‘, dieser Verwendungsaspekt von *Technik* ist im Bezugsrahmen des Paronymwörterbuchs folglich als „individueller Verwendungsaspekt ohne Ähnlichkeit“ einzuordnen.

Die Kookkurrenzanalyse eignet sich hervorragend für die korpusgestützte lexikografische Analyse von Einzelwörtern, gleich, ob man diese wie in *lexiko* möglichst umfassend beschreibt oder deren Beschreibung wie im Paronymwörterbuch möglichst knapp auf

einen Vergleich hin zuspitzen will. Ein Korpus bildet das Fundament, die Kookkurrenzanalyse die Methode der Analyse und Datengenerierung. Der Vergleich von *ellexiko* mit dem Paronymwörterbuch zeigt, wie unterschiedlich die Ergebnisse aus dem Zusammenspiel dieser beiden Bausteine lexikografischer Arbeit sein können. Dieser Umstand bedingt die Forderung nach Transparenz. Nur so kann erschlossen werden, wo welche Daten herkommen, ob und wie Daten in Beziehung miteinander stehen usw. Der Vergleich bzw. die Vergleichbarkeit bildet das oberste Strukturprinzip des Paronymwörterbuchs. Wie bereits gezeigt, stößt das Verfahren der Kookkurrenzanalyse damit an seine Grenzen: Eine kontrastiv ausgerichtete Untersuchung und Betrachtung zweier Suchlemmata (Paronyme) ist damit nicht möglich, sondern lediglich die eines isolierten Suchlemmas. Es ist daher Aufgabe der Lexikograf/innen, die Ergebnisse der zwei oder mehr gesondert für jedes Paronym durchgeführten Kookkurrenzanalysen in einem redaktionellen Prozess einem Vergleich zu unterziehen, Ähnlichkeiten und Unterschiede herauszuarbeiten und die entsprechenden Daten in die Artikelstruktur des Wörterbuchs einzupflegen. Die Komplexität dieser Aufgabe steigt mit zunehmender Anzahl von Paronymen bzw. Verwendungsaspekten. Daher werden als methodische Basis neben der Kookkurrenzanalyse auch andere Verfahren angewandt, die sich im Idealfall gegenseitig ergänzen (vgl. Teichmann in diesem Heft).

Der Anteil redaktioneller Arbeitsschritte auf dem Weg zur Kodierung des Wörterbucheintrags ist also beim Paronymwörterbuch nicht unerheblich. Es ist daher nur folgerichtig, die Menge der im Wörterbuch dokumentierten Informationen so gering wie möglich und so umfangreich wie nötig zu halten – mit stetem Blick auf den Vergleich. Nachvollziehbarkeit ist oberstes Gebot, sollen die Nutzer/innen die Daten auch im intendierten Sinn dekodieren, d. h. angemessen interpretieren können – umso mehr im Falle eines Vergleichs, was nun zum Ausgangspunkt der angestellten methodischen Überlegungen zurückführt. Kommen wir abschließend noch zur Dynamik.

Ähnlich klingende Wörter, die inhaltlich partielle Übereinstimmungen besitzen, führen zu Unsicherheiten, mitunter Verwechslungen und vielfach zu einem Nachschlagebedürfnis mit dem Ziel, mehr über deren angemessene Verwendung zu erfahren. Es gilt, Antworten zu finden, was speziell im für den Gebrauchswandel sehr offenen, dynamischen Bereich der Paronyme nicht immer ganz leicht ist. Neben sprachlich motivierten Faktoren scheinen in diesem Bereich vielfach auch andere, etwa kognitive Einflüsse eine Rolle zu spielen. Paronymie ist ein komplexes Phänomen, deutlich komplexer als man zunächst meinen mag. Dem sollte eine lexikografische Annäherung Rechnung tragen, was die Methodik, die Inhalte und die Präsentation betrifft. Die oben vorgenommene Gegenüberstellung zweier semantisch ähnlicher Verwendungsaspekte von *Technologie* und *Technik* demonstrierte, dass es nicht genügt, vergleichsrelevante Informationen auszuwählen, zu gruppieren und so zu formulieren, dass Gemeinsamkeiten und Unterschiede ‚irgendwie‘ zutage treten: In der Präsentation laufen die Fäden zusammen. Bislang war meist von ‚einem‘ Vergleich die Rede, und ein Vergleich bedarf u. a. eines tertium comparationis. Das jedoch kann sich schnell ändern. Daher ist es sinnvoll, die Darstellung des Paronymwörterbuchs nicht auf einen Vergleich zu beschränken, sondern mehrere Vergleichsmöglichkeiten zu implementieren und diese dynamisch zu gestalten.

Je nach dem Erkenntnisinteresse der Nutzer/innen können die Vergleichsparameter bei der Betrachtung von Paronymen variieren: Man kann sich einem Vergleich von Paronymen über die Häufigkeit ihrer Verwendungsaspekte nähern, sein Augenmerk auf die Gemein-

samkeiten legen oder aber die Unterschiede genauer untersuchen wollen bzw. all dies im wechselnden Modus. Die Konzeption des Paronymwörterbuchs lässt genau all dies zu. Nutzer/innen können wechselnde Vergleichsparameter im Wörterbuch auswählen, worauf sich die Präsentation der lexikografischen Inhalte entsprechend ändert. Die Auswahlmöglichkeiten haben zur Folge, dass unterschiedlich organisierte Ansichten bereits auf dem Bildschirm generiert werden, und nicht erst mühsam im Kopf der Nutzer/innen, die sich dadurch ganz auf die Untersuchung der selbst gewählten Vergleichsinhalte konzentrieren können. Die Änderung der ‚Vergleichsansichten‘ kann mitverfolgt werden: Diese Dynamik soll jedoch nicht ‚bloß ein Wenig Bewegung auf den Bildschirm bringen‘ oder hübsche Eyecatcher erzeugen. Die verfolgbaren Übergänge (‚transitions‘) von einer Ansicht in eine andere besitzen Funktionen, deren offenkundigste ist, die unterschiedlichen Ergebnisse bei Änderung eines Vergleichsparameters zu visualisieren und damit nachverfolgbar zu machen. Dies ist kaum möglich, wenn unterschiedliche Ansichten mittels „klick“ einfach ‚aus dem Hut gezaubert‘ werden; Zusammenhänge oder Orientierungsanker können so verlorengehen (vgl. zu Aspekten der Dynamik auch Storjohann in diesem Heft).

### 3. Resümee

Der Paronymstatus des Wortpaares ist nicht ‚gegeben‘, sondern muss erst eingehend anhand seiner Verwendung(en) und damit seiner Kollokatoren untersucht werden: Wortbildungsregeln allein lassen nur in sehr begrenztem Maß Rückschlüsse auf das semantisch relationale Verhältnis der Paare zu (vgl. Reckenthäler in diesem Heft), außerdem gibt es ganz unterschiedliche Typen von Paronymen (vgl. Mell/Schnörch/Storjohann in diesem Heft). Die lexikografische Arbeit zur vergleichenden Beschreibung von Ausdrücken in paronymer Relation beginnt daher jedes Mal von ‚vorne‘, also beim Korpus. Mit der Kookkurrenzanalyse steht ein probates Analyseverfahren zur Corpus-driven-Analyse signifikanter Phänomene im Sprachgebrauch zur Verfügung (siehe auch Perkuhn in diesem Heft). Von Vorteil ist, dass dieses im Gegensatz zu den anderen in diesem Band vorgestellten SOM-basierten Analyseverfahren (siehe Teichmann in diesem Heft) auf einem eigens für die Belange des Paronymwörterbuchs zusammengestellten Untersuchungskorpus basiert. Das größte Manko der Kookkurrenzanalyse ist, dass das automatisierte Verfahren stets nur für ein einzelnes Stichwort angewandt werden kann. Der Vergleich muss mit meist größerem Aufwand von den Lexikograf/innen redaktionell bewerkstelligt werden. Darin liegt jedoch die Chance, dass die Wörterbuchinhalte sehr fundiert, ausgewogen, deskriptiv und auf den Vergleich zugeschnitten erarbeitet werden können. Auf diesem Weg lässt sich ein hoher Grad an Zuverlässigkeit erreichen, der nicht von allen Nachschlageressourcen gewährleistet wird. So heißt es beispielsweise in einer dpa-Meldung unter der Überschrift „Erfahrener Wikipedia-Autor warnt vor Artikeln des Online-Lexikons“:

Statt Informationen blind zu vertrauen, sollten Nutzer [...] genau überprüfen, ob diese aus zuverlässigen Quellen stammen [...]. Das sei nötig, weil für die Qualitätskontrolle der Artikel Laien verantwortlich sind. (Augsburger Allgemeine online, 5. 11. 2017)

Ob die Qualität und schiere Menge der Daten, die man bei anderen häufig frequentierten rein algorithmisch arbeitenden Auskunftsquellen (z. B. Suchmaschinen wie Google) erhält, in ihrer Unverbindlichkeit und Masse wirklich weiterhelfen, mag an dieser Stelle jede(r) für sich beantworten. Der fertige Wörterbuchartikel des Paronymwörterbuchs soll sowohl den Inhalten als auch der Präsentation nach möglichst auf den ersten Blick die in diesem Zusammenhang vielleicht häufigste Frage von Nutzer/innen zuverlässig beant-

worten: „Kann ich ein Paronym in diesem Kontext verwenden, oder ist das ‚falsch‘?“ (vgl. z. B. auch Storjohann 2016). Bei der Beantwortung ist an erster Stelle Zuverlässigkeit gefragt (vgl. Müller-Spitzer (Hg.) 2014), die andere Quellen häufig nicht bieten (siehe oben). Darüber hinaus möchte das Wörterbuch aber auch gesicherte Auskunft auf andere Fragen im Zusammenhang mit dem Gebrauch bzw. Vergleich von Paronymen geben, im Idealfall schnell und mit wenigen Klicks, Touches oder Swipes.

## Literatur

Belica, Cyril (1995): Statistische Kollokationsanalyse und -clustering. Korpuslinguistische Analyseverfahren. Mannheim: Institut für Deutsche Sprache. <http://corpora.ids-mannheim.de> (Stand: 17.6.2014).

*elexiko*: [www.elexiko.de](http://www.elexiko.de) (Stand: 6.11.2017).

Haß, Ulrike (Hg.) (2005): Grundfragen der elektronischen Lexikographie. *elexiko* – das Online-Informationssystem zum deutschen Wortschatz. Berlin/New York: De Gruyter. (= Schriften des Instituts für Deutsche Sprache 12).

Müller-Spitzer, Carolin (Hg.) (2014): Using online dictionaries. Berlin: De Gruyter. (= Lexicographica. Series Maior 145).

Paronyme – Dynamisch im Kontrast. [www.owid.de/plus](http://www.owid.de/plus) (Stand: 6.11.2017).

Paronymkorpus. [www1.ids-mannheim.de/lexik/paronymwoerterbuch/dasparonymkorpus.html](http://www1.ids-mannheim.de/lexik/paronymwoerterbuch/dasparonymkorpus.html) (Stand: 6.11.2017).

Perkuhn, Rainer/Keibel, Holger/Kupietz, Mark (2012): Korpuslinguistik. Paderborn: Fink.

Schnörch, Ulrich (2005): Der Autoren-Arbeitsplatz: Ein *elexiko*-Wortartikel entsteht. In: Haß (Hg.), S. 105–130.

Schnörch, Ulrich (2015): Vom Korpus ins Wörterbuch. Die Ermittlung und Auswertung von Sprachdaten für *elexiko*, ein korpusgestütztes Online-Wörterbuch zur deutschen Gegenwartssprache. In: Kilian, Jörg/Eckhoff, Jan (Hg.): Deutscher Wortschatz – beschreiben, lernen, lehren. Beiträge zur Wortschatzarbeit in Wissenschaft, Sprachunterricht, Gesellschaft. Frankfurt a. M. u. a.: Lang. S. 39–54. (= Germanistik – Didaktik – Unterricht 13).

Spillner, Bernd (1997): Stilvergleich von Mehrfachübersetzungen ins Deutsche (ausgehend von Texten Dantes und Rimbauds). In: Fix, Ulla/Wellmann, Hans (Hg.): Stile, Stilprägungen, Stilgeschichte. Über Epochen-, Gattungs- und Autorenstile. Sprachliche Analysen und didaktische Aspekte. Heidelberg: Winter. S. 207–210. (= Sprache – Literatur und Geschichte 15).

Storjohann, Petra (2016): Vom Interesse am Gebrauch von Paronymen zur Notwendigkeit eines dynamischen Wörterbuchs. In: Sprachreport 32, 4. S. 32–43.

Dr. Ulrich Schnörch  
Institut für Deutsche Sprache Mannheim  
R5, 6–13  
68161 Mannheim  
E-Mail: [schnoerch@ids-mannheim.de](mailto:schnoerch@ids-mannheim.de)