

Das wissenschaftlich- religionspädagogische Lexikon im Internet

(WiReLex)

Jahrgang 2016

Korpusanalyse

Stefan Altmeyer

erstellt: Februar 2019

Permanenter Link zum Artikel:

<http://www.bibelwissenschaft.de/stichwort/200609/>



DEUTSCHE
BIBEL
GESELLSCHAFT

Korpusanalyse

Stefan Altmeyer

1. Begriff und Anliegen

Eine *Korpusanalyse* ist ein Verfahren der empirisch arbeitenden Sprachwissenschaft mit dem Ziel, einen konkret umrissenen Sprachgebrauch methodisch abgesichert zu beschreiben und Hypothesen bezüglich der Merkmale dieser Sprache zu bilden oder zu prüfen. Ihr empirisches Material ist das *Korpus*, eine nach expliziten Kriterien erstellte, digital verfügbare Zusammenstellung von authentischen Beispielen geschriebener oder transkribierter gesprochener Sprache. Mittels quantitativer und qualitativer Methoden werden Sprachgebrauchsmuster identifiziert und von ihrem Gebrauchskontext her interpretiert. Die verwendeten Methoden dienen im Unterschied zur sozialempririschen Forschung (→ [Empirie](#)) nicht der „Erhebung und Interpretation von Erfahrungswissen“ (Ziebertz, 2001, 1746), sondern haben „Sprachgebrauchswissen“ (Bubenhofner, 2009, 43) zum Gegenstand, sodass differenzierend von „sprachempirischer Forschung“ (Altmeyer, 2011, 129) gesprochen werden kann. Insofern jedoch beide Formen des Wissens miteinander in Verbindung stehen – Erfahrungswissen liegt meist in Form von → [Sprache](#) vor, und Sprachgebrauch verweist umgekehrt auf Erfahrungen (→ [Erfahrung](#)) in je konkreten Sprechkontexten –, ist von einer wechselseitigen Verwiesenheit beider Ansätze auszugehen.

Beispiel: Zur Frage nach den Gottesvorstellungen von Schülerinnen und Schülern (→ [Gott](#)) gibt es viele sozialemprirische Studien, die durch Interviews, Befragungen, Beobachtungen etc. Erfahrungswissen erhoben und Theorien entwickelt haben. Ein sprachempirisches Vorgehen kann diese Forschungen ergänzen, indem es den Fokus auf den Sprachgebrauch der Schülerinnen und Schüler lenkt. So hat etwa Altmeyer (2011) über 2.000 kurze, im Religionsunterricht entstandene Texte mit einem korpuslinguistischen Setting untersucht und stellte dabei u.a. fest, dass die Jugendlichen ihre Gottesvorstellungen häufig durch apersonale, positive Erfahrungsbegriffe wie Hoffnung, Kraft, Halt und Geborgenheit zum Ausdruck bringen.

Eine Korpusanalyse ist also für empirische Projekte in der Religionspädagogik immer dort von Interesse, wo eine Untersuchung umfangreicheren Sprachmaterials (Texte; offene Items in Fragebögen (→ [Fragebogen](#));

gesellschaftliche, wissenschaftliche oder kirchliche Diskurse; Online-Kommentare etc.) angezielt ist. Ihre Stärke besteht darin, mit geeigneten Methoden Merkmale und Strukturen an der Sprachoberfläche sichtbar zu machen, auf deren Basis neben linguistischer Theoriebildung auch Zugänge zu den zugrundeliegenden Bedeutungssystemen möglich werden. Gerade Letzteres ist für die religionspädagogische Rezeption relevant (gute Einführung: Scott/Tribble, 2006; deutschsprachige Lehrbücher: Lemnitzer/Zinsmeister, 2015; Perkuhn/Keibel/Kupietz, 2012; Scherer, 2014; umfassende Überblicke: Lüdeling/Kytö, 2008f.; O'Keeffe/McCarthy, 2010).

2. Hintergründe

2.1. Korpuslinguistik: Theoretische Grundlagen

Die systematische Untersuchung von großen Textsammlungen beginnt nicht erst mit der Etablierung der modernen Korpuslinguistik ab Ende der 1960er Jahre, sondern hat eine jahrhundertelange Tradition – nicht zuletzt auch in der biblischen Exegese (Lüdeling/Kytö, 2008f., Bd. 1, 1-153). Die gegenwärtige große Bedeutung korpuslinguistischer Verfahren verdankt sich jedoch vor allem den umfassenden Möglichkeiten digitaler Texterfassung und -erschließung, die inzwischen auch im Kontext der sogenannten *digital humanities* Verbreitung finden. Die immer einfachere und umfassendere Verfügbarkeit von digitalisierten Sprachdaten wirft jedoch mit Nachdruck die Frage auf, wie überhaupt aus solchen Daten valides und interpretierbares → [Wissen](#) gewonnen werden kann.

Die Antwort der Korpuslinguistik besteht in einem pragmatisch gewendeten Sprachmodell, das im sogenannten britischen Kontextualismus (Lemnitzer/Zinsmeister, 2015, 30-33; Hintergrund: Krämer, 2006) theoretisch verortet ist. Dessen zentrale Prämisse lautet, dass sich die Bedeutung einer sprachlichen Äußerung aus der Analyse des Sprachgebrauchs erschließt, wie er sich im Kontext der untersuchten Äußerung niederschlägt. Nicht das Wissen eines Sprechers (Kompetenz) ist der entscheidende Zugang zum sprachlichen Verstehen, sondern die Beziehung zwischen Form, Inhalt und Kontext konkreter Äußerungen. Um also von bloßen Sprachdaten zu theoretischem Wissen zu gelangen, sind sprachliche Äußerungen „als Funktionen des sprachlichen und nicht-sprachlichen Kontextes zu erklären, in dem diese Äußerungen stehen“ (Lemnitzer/Zinsmeister, 2015, 30). Korpuslinguistik betrachtet daher sprachliche Äußerungen in unterschiedlichen Kontexten, die vom unmittelbaren Wortumfeld bis hin zum sozialen Kontext reichen. Ganz grundlegend ausgedrückt, geht es dabei um empirisches Wissen darüber, wie im Sprachgebrauch Bedeutung erzeugt und transportiert wird (Sinclair, 2007).

2.2. Methodologische Paradigmen

Drei paradigmatische Verfahrensweisen lassen sich unterscheiden, die ein je eigenes Erkenntnisinteresse verfolgen und auf einem unterschiedlichen Einsatz quantitativer und qualitativer Methoden beruhen (Mahlberg, 2005, 16-39; Tognini-Bonelli, 2001). Als quantitative Methoden gelten in der Korpuslinguistik solche, die wesentlich auf Häufigkeitsanalysen beruhen und daraus bestimmte Maßzahlen berechnen, mit deren Hilfe sich Korpora charakterisieren und vergleichen lassen (Biber/Jones, 2009). Qualitative Verfahren hingegen sind daran interessiert, einzelne Phänomene des Sprachgebrauchs zu entdecken und durch Kontextuntersuchungen zu beschreiben, zu klassifizieren und zu interpretieren, wobei aber auch statistische Verfahren eine Rolle spielen können (Wynne, 2008). Anders als im sozialempririschen Bereich verweist die Unterscheidung quantitativ/qualitativ hier also nicht auf unterschiedliche Arten von Daten, sondern auf einen unterschiedlichen Auflösungsgrad der empirischen Analyse (→ [Qualitative Sozialforschung in der Religionspädagogik](#)).

Als erstes methodisches Paradigma lässt sich das *korpusgestützte Modell* benennen, das tendenziell deduktiv vorgeht. Korpora kommen hier zum Einsatz, um vorab formulierte Theorien zu überprüfen, was mittels quantitativer wie qualitativer Verfahren geschehen kann.

Das zweite Paradigma, das *korpusbasierte Modell*, beschreibt ein induktives Erkenntnisinteresse. In der Korpusanalyse werden Texte exhaustiv erschlossen, das heißt ohne dass von vorneherein Daten aufgrund theoretischer Hypothesen ausgeschlossen würden. Mit Schwerpunkt auf quantitativen Methoden werden etwa Häufigkeiten erhoben oder charakteristische Kennzahlen berechnet, die eine objektivierte Basis der Korpusbeschreibung bieten. Auch computerisierte Verfahren zur automatischen Informationsgewinnung aus umfangreichen Datensätzen (Data/Text Mining) lassen sich hier einordnen.

Im *korpus-theoretischen Modell* bildet der Sprachgebrauch im Korpus ebenfalls den Ausgangspunkt. Allerdings werden hier im Sinne eines abduktiven Vorgehens empirische Befunde und Theoriebildung abgewechselt (Mahlberg, 2005, 179-194). Kein gefundenes Sprachmerkmal bleibt ohne Interpretation, sondern führt zu einer Hypothesenbildung, die wiederum am Sprachgebrauch überprüft wird und so weiter. Hierbei werden quantitative und qualitative Methoden miteinander verknüpft mit dem Ziel der Konstruktion plausibler Theoriemodelle.

Beispiel: Der Sprachwissenschaftler Wolfgang Teubert (2007) hat Dokumente der Katholischen Soziallehre von *Rerum novarum* (1891) bis *Centesimus annos* (1991) untersucht und dabei die Bedeutung von Grundbegriffen wie Eigentum und Arbeit in den Blick genommen. In seiner Analyse verfolgt er die Entwicklung des Begriffsgebrauchs innerhalb des Korpus und untersucht, welche Diskurse von außerhalb aufgegriffen und verarbeitet werden. Damit

wählt er einen korpuslinguistischen Ansatz, der Funde im Korpus sukzessiv interpretiert durch Bezug auf weitere Kontexte, die Spuren im Korpus hinterlassen haben. Eine korpusbasierte Herangehensweise würde sich im Unterschied dazu auf den Sprachgebrauch im Korpus beschränken, während eine korpusgestützte Analyse dazu geeignet wäre, bereits vorab formulierte Hypothesen etwa über das Verständnis von Eigentum in der Katholischen Soziallehre am Korpus zu überprüfen.

3. Anwendungen

Das Spektrum korpuslinguistischer Fragestellungen und Analysemethoden ist nahezu unüberschaubar. Grundvoraussetzung für die Anwendbarkeit ist lediglich eine Forschungsfrage, die sich durch die Analyse eines definierten Sprachgebrauchs beantworten lässt. Im Folgenden wird ein knapper, an Beispielen orientierter Einblick in grundlegende Anwendungsmöglichkeiten gegeben, die insbesondere für religionspädagogische Forschungen interessant sind. Eine Korpusanalyse besteht dabei im Wesentlichen aus drei Aufgaben: der Erschließung von Daten (Korpuserstellung), der Analyse von Sprachgebrauchsmustern sowie der Bildung von interpretierenden Hypothesen.

3.1. Erschließung von Daten

Die Zuverlässigkeit der Ergebnisse einer Korpusanalyse ist abhängig von Auswahl und Aufbereitung der Korpus Texte. Als die drei wichtigsten Kriterien werden in der Regel Authentizität, Repräsentativität und Computerisierung genannt (Mukherjee, 2009, 21-23; Sinclair, 2005). Authentizität bedeutet, dass die Sprachbeispiele auf linguistisch unreflektierten Sprachgebrauch zurückgehen. Repräsentativität ist hier nicht identisch mit der Auswahl einer repräsentativen Stichprobe im Sinne der empirischen Sozialforschung, da die Grundgesamtheit aller möglichen sprachlichen Äußerungen grundsätzlich nicht bestimmt werden kann (Biber, 2007). Gemeint ist vielmehr die graduelle Frage, in welchem Maße das untersuchte Korpus eine für die Forschungsfrage inhaltlich und formal ausgewogene und aussagekräftige Auswahl an Sprachbeispielen darstellt.

Das Kriterium der Computerisierung besagt, dass Sprachdaten (schriftlich wie mündlich) in eine computerlesbare Form zu bringen sind, wofür sich bestimmte Standards entwickelt haben (Lemnitzer/Zinsmeister, 2015, 39-56). Am wichtigsten ist hierbei die Unterscheidung zwischen Primär- und Metadaten. Während die in einem Korpus enthaltenen Texte als Primärdaten bezeichnet werden, versteht man unter Metadaten solche Informationen, die zur weiteren Erschließung dieser Texte hinzugefügt wurden und entweder außertextliche Kontextdaten (Autor, Quelle, Ort und Zeit der Entstehung etc.) oder auch umfassende linguistische Informationen (Wortart, Flexionsform etc.) umfassen

können. Während erstere für eine religionspädagogische Korpusanalyse unerlässlich sind, dürfte die sogenannte linguistische Annotation schon aus pragmatischen Erwägungen in den meisten Fällen verzichtbar sein. Grundsätzlich zu beachten ist, dass bereits die Erstellung eines analysierbaren eigenen Korpus je nach Art der Sprachdaten einen nicht unerheblichen Aufwand darstellt.

Beispiel: Eine Forschungsarbeit interessiert sich dafür, was die wichtigsten Themen in der religionspädagogischen Diskussion in Deutschland der letzten fünf Jahre gewesen sind. Publierte Texte sind für eine wissenschaftliche Disziplin von zentraler Bedeutung, sodass die Korpusanalyse eine passende Herangehensweise darstellt. Da ein entsprechendes Korpus nicht schon vorliegt, müssen selbst Texte zu einem Korpus zusammengestellt werden. Während das Kriterium der Authentizität relativ unproblematisch ist, stellt Repräsentativität angesichts der Vielzahl und Diversität der Veröffentlichungen eine große Hürde dar. Eine pragmatische Möglichkeit ist zu fragen, welche Texte leicht zugänglich sind. Dies trifft beispielsweise für die Artikel des WiReLex zu, die schnell zu einem lesbaren Korpus kompiliert werden können. Allerdings kann mit diesem Korpus nicht der gesamte religionspädagogische Diskurs repräsentiert werden, wenngleich angesichts von 368 Artikeln (Stand 08/2018) doch schon weitreichende Aussagen möglich scheinen. Die folgenden Beispiele greifen auf dieses Korpus unter dem Namen WiReLex2018 zurück. Alle Daten der hier vorgestellten Analysen stehen [hier](#) zur Verfügung.

3.2. Analyse von Sprachgebrauchsmustern

Für die Korpusanalyse gibt es speziell entwickelte Softwarelösungen, die häufig auf bestimmte Fragestellungen oder vorhandene große Korpora zugeschnitten sind. Als umfassendes, für alle hier vorgestellten Anwendungen nutzbares Paket, sei auf das von Mike Scott entwickelte Analysetool WordSmith (Scott, 2016) und das begleitende [Handbuch](#) verwiesen.

3.2.1. Typische Wörter

Die kleinste und zugleich grundlegendste Analyseeinheit ist das einzelne Wort (type) und dessen Vorkommen an bestimmten Stellen im Korpus (token). Die einfachste quantitative Analyse ist daher die Häufigkeitsanalyse, die danach fragt, wie häufig ein Wort vorkommt und wie sich dieses Vorkommen in einem Korpus verteilt (Scott/Tribble, 2006, 11-32). Ergebnisse von Häufigkeitsanalysen lassen sich in Form von Tabellen oder auch als Visualisierung (z.B. Wortwolke, Verteilungsplot) darstellen und geben einen ersten Einblick in das sprachliche Grundgerüst eines Korpus.

Beispiel: Das Ergebnis einer Häufigkeitsanalyse in WiReLex2018 liefert

zunächst die typischerweise häufigsten Wörter der deutschen Sprache (Artikel, Konjunktionen, Präpositionen), doch sobald in der Liste Nomen auftauchen, wird deutlich, was die Korpus Texte inhaltlich umtreibt: Religionsunterricht, Menschen, Bildung, Gott, Religion, Schülerinnen, Kirche besitzen die häufigsten Nennungen. Während Religionsunterricht in 76% aller Texte vorkommt, wird beispielsweise Erwachsenenbildung nur in 17% angesprochen. Schaut man nur die Literaturverzeichnisse an, sieht man z.B., dass Friedrich Schweitzer mit Abstand am häufigsten zitiert wird, und zwar in 29% aller Artikel.

Ein differenziertes und aussagekräftiges quantitatives Verfahren ist die Schlüsselwortanalyse (Scott/Tribble, 2006, 55-72; Hintergrund: Bondi/Scott, 2010; Gabrielatos, 2018). Dabei werden nicht einfach absolute Häufigkeiten gezählt, sondern man fragt, ob das Vorkommen eines Wortes höher ist, als man es normalerweise erwarten würde. Unter einem Schlüsselwort versteht man dabei ein Wort, das im untersuchten Korpus im Vergleich zu einer Norm (Referenzkorpus) signifikant häufig (oder selten) gebraucht wird, wobei ein statistisches Maß für diese Signifikanz angegeben werden kann. Eine Schlüsselwortanalyse ist geeignet, erste wesentliche Muster im Sprachgebrauch aufzudecken, und bildet einen oftmals geeigneten Ausgangspunkt für Hypothesenbildung und weitere Detailuntersuchungen.

Beispiel: Eine Schlüsselwortanalyse für WiReLex2018 bringt aufgrund der Korpusgröße sehr viele Ergebnisse, die am besten zunächst formal oder inhaltlich geordnet werden. So finden sich unter den ersten 200 Schlüsselwörtern solche, die z.B. die meist behandelten Inhalte und Lernorte oder die wichtigsten Grundbegriffe erkennen lassen. Eine andere Klasse zeigt, welche Konzepte in der Religionspädagogik aktuell zentral sind: → [Elementarisierung](#), Kinder-/Jugendtheologie (→ [Kindertheologie](#); → [Jugendtheologie](#)) und → [Korrelation](#). Man entdeckt auch Schlüsselwörter, deren prominente Position überrascht, wie z.B. Sünde. Ein genauerer Blick zeigt, dass sich die Funde auf relativ wenige Texte verteilen, doch kommt das Wort immerhin noch in 9% aller Artikel vor!

Die Ergebnisse einer Schlüsselwortanalyse hängen per Definition von der gewählten Vergleichsnorm ab, die entsprechend der Fragestellung zu wählen ist. Es gibt große Korpora, die als Referenzkorpus die deutsche Gegenwartssprache ausgewogen zu repräsentieren versuchen (Lemnitzer/Zinsmeister, 2015, 142-148; z.B. mit umfassenden Suchmöglichkeiten: [DWDS](#)). Die Verwendung eines solchen Korpus ist meistens hilfreich, um typische Themen im Korpus aufzudecken. Mittels Auswahl des Referenzkorpus können aber auch gezielt kontrastive Fragestellungen adressiert werden. So kann man etwa Teilmengen eines größeren Korpus (sogenannte Teilkorpora) miteinander vergleichen. Ein

Teilkorpus kann nach sprachlichen oder kontextuellen Kriterien gebildet werden, z.B. alle Texte, die ein bestimmtes Wort enthalten, oder alle Texte männlicher Autoren. Zu beachten ist die größtmögliche Vergleichbarkeit der gebildeten Teilkorpora wie auch die Tatsache, dass durch die Teilmengenbildung unter Umständen schon Unterschiede unterstellt werden, die eigentlich erst noch untersucht werden sollen.

Beispiel: Vorherige Analysen mit WiReLex2018 hatten Friedrich Schweitzer als meistzitierten Autor ergeben. Man kann das Gesamtkorpus in Texte mit bzw. ohne Referenz auf diesen Autor aufteilen und die beiden Teilkorpora miteinander vergleichen, um die Themen herauszufinden, die mit ihm verbunden bzw. weniger verbunden werden. Das Ergebnis ist eindeutig: Die Schlüsselwortanalyse ergibt gut 20 Themen, zu denen der Autor bisher nicht konsultiert wird, darunter: → [Kirchengeschichte](#), Paulus, → [Musik](#), → [Medien](#) und der Papst. Eine schöne Liste, um das Thema für ein nächstes Buch zu finden!

Arbeitet man mit einem Korpus, das aus vielen Einzeltexten zusammengesetzt ist, kann eine zweistufige Schlüsselwortanalyse interessant sein, durch die Schlüsselwörter mehrerer Texte miteinander in Beziehung gesetzt werden. Es wird berechnet, welche Schlüsselwörter für mehrere Texte typisch sind und in welchem Umfang dies zutreffend ist (Scott/Tribble, 2006, 73-88). Damit lässt sich z.B. feststellen, ob bestimmte interessante Textgruppen Schlüsselwörter gemeinsam haben oder gerade nicht.

Beispiel: Eine zweistufige Schlüsselwortanalyse kann z.B. aufzeigen, dass der Islam in 12 Texten (3%) von WiReLex2018 markant thematisiert wird und damit häufiger als Christentum (10 Texte) und seltener als Kirche (41 Texte – genauer: diese Wörter sind in einer entsprechenden Zahl von Texten ein Schlüsselwort). In den Texten zum Islam tauchen darüber hinaus häufig noch Muslime, Dialog, Begegnung und das Judentum als Schlüsselwörter auf, was deutlich aufzeigt, in welchem Kontext und mit welchem Schwerpunkt der Islam thematisiert wird.

3.2.2. Wortverbindungen

Bieten die bisherigen Verfahren wie in einem Satellitenbild einen Makroblick auf die Sprachgebrauchsmuster, so gibt es andere, um dieses weite Bild in ausgewählten Ausschnitten heranzuzoomen. Häufig verwendet man dabei eine sogenannte Kollokationsanalyse (Scott/Tribble, 2006, 33-35; Hintergrund: Evert, 2009). Hierzu werden der unmittelbare Kontext (Wortumfeld) eines Suchworts und damit seine mikrologischen Gebrauchsmuster untersucht. Das gemeinsame Vorkommen zweier Wörter nennt man Kollokation. In kleineren Korpora lassen

sich diese Verbindungen am Einzelfall analysieren und interpretieren, während sich in größeren Korpora signifikante Kollokatoren eines Suchwortes durch statistische Funktionen extrahieren lassen. Jeweils geht es darum, typische Wortkombinationen sichtbar zu machen, die auf verschiedene Bedeutungsmuster hinweisen können. Dabei kann eine Visualisierung in Form von Strukturgraphen hilfreich sein.

Beispiel: Das Wort Gott ist in WiReLex2018 ein hoch frequentes Schlüsselwort. Es lässt sich fragen, auf welche Weise in religionspädagogischen Artikeln auf Gott Bezug genommen wird. Eine Visualisierung der häufigsten Kollokatoren (Abb. 1) zeigt, dass dabei ein bestimmtes inhaltliches, nämlich vor allem ein beziehungsorientiertes Gottesbild aufgerufen wird.

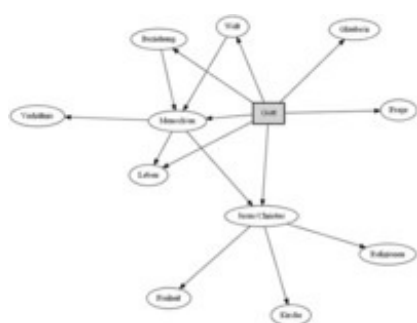


Abb. 2 Kollokationen des Schlüsselwortes Gott in WiReLex2018

Die Suche nach Verbindungen zwischen Wörtern lässt sich auch mit der Schlüsselwortanalyse kombinieren. Die Frage ist dann, ob auch Schlüsselwörter häufiger gemeinsam vorkommen, als der Zufall es erwarten ließe. Signifikant miteinander verbundene Schlüsselwörter heißen Associates, assoziierte Schlüsselwörter (Scott/Tribble, 2006, 85). Zusätzlich kann man erforschen, ob Schlüsselwörter nicht nur im selben Text, sondern im engeren Wortumfeld stehen (und also Kollokatoren sind).

3.2.3. Kontexte

Die größte Detailschärfe gewinnt die Korpusanalyse in unmittelbaren Kontextuntersuchungen. Ausgangspunkt dazu ist in der Regel die Suche nach bestimmten Wörtern, die sich entweder durch die Forschungsfrage oder vorangegangene Analysen ergeben. Das elementare Werkzeug qualitativer Korpusanalyse ist die Konkordanz (Scott/Tribble, 2006, 33-53; Hintergrund: Wynne, 2008). Dabei handelt es sich um eine aufbereitete Liste aller Fundstellen eines Suchwortes innerhalb seines unmittelbaren Kontextes benachbarter Wörter. Die Konkordanz ist Ausgangspunkt und Hilfsmittel für alle Analysen, die Mikrokontexte in den Blick nehmen.

Beispiel: Eine Detailanalyse interessiert sich für das Profil des Grundbegriffs Dialog in WiReLex2018. Schon eine erste Sichtung der Konkordanz von 322 Fundstellen aus 123 Texten lässt erkennen, dass hierbei unterschiedliche Spezifizierungen zu beachten sind. Macht es einen Unterschied, ob vom

christlich-islamischen, christlich-jüdischen oder einem interreligiösen Dialog gesprochen wird? Was bedeuten ein kritischer, ein theologischer, ein religiöser Dialog?

Während in der Konkordanz vor allem das unmittelbare Wortumfeld untersucht wird, kann der Kontext auch weiter gefasst werden, was sich in Form konzentrischer Kreise beschreiben lässt (vgl. Abb. 2; Altmeyer, 2011, 136).



Abb. 4 Dimensionsmodell des sprachlichen Kontextes

Eine wichtige Rolle spielt hierbei der Übergang vom textlichen Kontext (Kotext) zum außertextlichen Kontext, welcher die Rahmenbedingungen einer Sprachhandlung (Situation, Kultur bis hin zum Rezeptionszusammenhang) beschreibt (Scott/Tribble, 2006, 3-10). Gerade an diesen Schnittstellen wird deutlich, wo Übergänge zu anderen sprachanalytischen aber auch

sozialempirischen Herangehensweisen bestehen: Eine Korpusanalyse lässt sich je nach Fragestellung mit diskursanalytischen (→ [Diskursanalyse](#)), hermeneutischen (→ [Hermeneutik](#)) wie auch sozialempirischen Herangehensweisen (→ [Forschungsmethoden, religionspädagogische](#)) triangulieren (Baker, 2006, 2010; Mahlberg, 2014).

Beispiel: Altmeyer u.a. (2015) haben in einer kulturell vergleichenden Studie untersucht, was Menschen in Deutschland und in den USA meinen, wenn sie sich als religiös oder spirituell bezeichnen. Als Teil eines umfangreichen Fragebogens ließen sie die Probanden auch freie Texte zu ihrem Verständnis von Religion und Spiritualität schreiben. Durch die Anlage der Untersuchung konnten Sprachgebrauch und andere sozialempirische Daten miteinander verknüpft werden. Es zeigte sich, dass Menschen je nach religiöser Selbsteinschätzung (als religiös, spirituell, weder noch oder beides) auch etwas anderes unter den Konzepten verstehen.

3.3. Hypothesenbildung

Die korpusanalytische Untersuchung eines bestimmten Sprachgebrauchs kommt weder ohne theoretische Vorannahmen aus, noch sind ihre Ergebnisse frei von Interpretationen (Sinclair, 2004, 9-23). Zudem ist es für religionspädagogische Forschungsvorhaben unabdingbar, die Befunde innerhalb des Verwendungszusammenhangs der eigenen Disziplin zu rekontextualisieren. Entsprechend ist die schrittweise Entwicklung einer im Sprachgebrauch gegründeten Theorie (→ [Grounded Theory](#)) zu empfehlen, wie sie auch im korpus-theoretischen Ansatz gefordert wird (Mahlberg, 2005, 179-194).

Entscheidender methodologischer Punkt ist in diesem Zusammenhang, dass die aus den Beobachtungsdaten hervorgehenden Verallgemeinerungen nachvollziehbar begründet werden. Dies kann durch die systematische Verwendung einer sogenannten Ausgangshypothese realisiert werden (Abb. 3). Diese dient dazu, die interpretierenden Vorannahmen bewusst zu machen, die mit einer Fragestellung an ein Korpus herangetragen werden. Nach den einzelnen beschreibenden Analyseschritten wird die Ausgangshypothese interpretierend weiterentwickelt bis hin zu einem theoretischen Modell, welches das systematisierte Sprachgebrauchswissen im Kontext religionspädagogischer Theoriebildung interpretiert. Dadurch wird der entscheidende Schritt zwischen sprachempirischer Beschreibung und theoretischer Interpretation transparent.

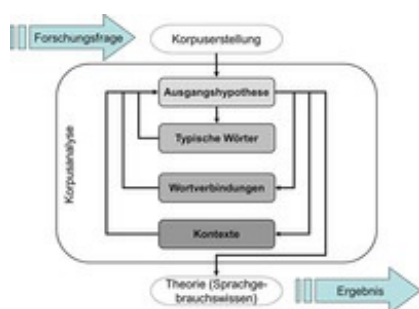


Abb. 6 Ablaufschema einer sprachempirischen Theoriebildung

4. Erträge

Die Anwendung korpusanalytischer Verfahren hat in der empirisch-religionspädagogischen Forschung bislang noch wenig Verbreitung gefunden, wenngleich Anwendungen und mögliche Erträge auf der Hand liegen (vgl. die [Liste](#) bislang vorliegender Studien). Überall dort, wo sprachlich gebundenes Wissen vorliegt, lassen sich entsprechende Methoden anwenden und mit dem bereits vertrauten

Vorgehen der empirischen Sozialwissenschaften verknüpfen. Zu denken wäre beispielsweise an eine systematische Erfassung der fachspezifischen Vorstellungen und Sprachkompetenzen von Schülerinnen und Schülern im Religionsunterricht in Abhängigkeit von möglichen Einflussfaktoren wie sozialer Lage, Herkunft, Familie, Schulart, Region etc. Ebenso ließen sich die charakteristischen Merkmale der überlieferten religiösen oder der aktuell in Kirche und Gesellschaft gebrauchten Sprache erheben und vergleichend danach fragen, auf welchen Feldern eine systematische religionspädagogische Förderung der religiösen Sprachkompetenz erfolgen müsste, wo diese anknüpfen könnte und wo allerdings auch mit typischen Schwierigkeiten zu rechnen wäre. Auch mediale und wissenschaftliche Diskurse sowie Texte in den sozialen Medien sind einer korpusanalytischen Erschließung gut zugänglich. Darüber hinaus bestehen Möglichkeiten, einfache korpusanalytische Methodenelemente (wie Häufigkeitsanalysen oder Wortwolken) auch unmittelbar mit Schülerinnen und Schülern im Unterricht einzusetzen (Altmeyer, 2013, 375f.), und zwar überall dort, wo ein reflektiertes Sprachbewusstsein aus inhaltlichen und fachlogischen Gründen unerlässlich erscheint. Insgesamt bietet die Korpusanalyse ein methodisches Potenzial, das in der religionspädagogischen Forschung und Praxis noch keineswegs ausgereizt scheint.

[Angaben zu Autor / Autorin finden Sie hier](#)

Empfohlene Zitierweise

Altmeyer, Stefan, Art. Korpusanalyse, in: Wissenschaftlich Religionspädagogisches Lexikon im Internet (www.wirelex.de), 2019

Literaturverzeichnis

- Altmeyer, Stefan, **Fremdsprache Religion? Sprachempirische Studien im Kontext religiöser Bildung, Praktische Theologie heute 114, Stuttgart 2011.**
- Altmeyer, Stefan, Die (religiöse) Sprache der Lernenden. Sprachempirische Zugänge zu einer großen Unbekannten, in: Becker-Mrotzek, Michael/Schramm, Karen/Thürmann, Eike (Hg. u.a.), Sprache im Fach. Sprachlichkeit und fachliches Lernen, Fachdidaktische Forschungen 3, Münster 2013, 365-379.
- Altmeyer, Stefan/Klein, Constantin/Keller, Barbara/Silver, Christopher/Hood, Ralph/Streib, Heinz, Subjective definitions of spirituality and religion. An exploratory study in Germany and the US, in: International Journal of Corpus Linguistics 20 (2015) 4, 526-552.
- Baker, Paul, Using corpora in discourse analysis, Continuum discourse series, London 2006.
- Baker, Paul, Sociolinguistics and corpus linguistics, Edinburgh sociolinguistics, Edinburgh 2010.
- Biber, Douglas, Representativeness in corpus design [1993], in: Teubert, Wolfgang/Krishnamurthy, Ramesh (Hg.), Corpus linguistics. Vol. 2, Critical concepts in linguistics, London 2007, 134-165.
- Biber, Douglas/Jones, James K., Art. Quantitative methods in corpus linguistics, in: Corpus Linguistics. An international Handbook II (2009), 1286-1304.
- Bondi, Marina/Scott, Mike (Hg.), Keyness in texts, Studies in corpus linguistics 41, Amsterdam 2010.
- Bubenhofer, Noah, Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse, Sprache und Wissen 4, Berlin 2009.
- Evert, Stefan, Art. Corpora and collocations, in: Corpus Linguistics. An international Handbook II (2009), 1212-1248.
- Gabrielatos, Costas, Keyness analysis. Nature, metrics and techniques, in: Taylor, Charlotte/Marchi, Anna (Hg.), Corpus approaches to discourse. A critical review, London 2018, 225-258.
- Krämer, Sybille, Sprache, Sprechakt, Kommunikation. Sprachtheoretische Positionen des 20. Jahrhunderts, Suhrkamp-Taschenbuch Wissenschaft 1521, Frankfurt 2006.
- Lemnitzer, Lothar/Zinsmeister, Heike, **Korpuslinguistik. Eine Einführung, Narr Studienbücher, Tübingen 3. Aufl. 2015.**
- Lüdeling, Anke/Kytö, Merja (Hg.), Corpus Linguistics. An international Handbook. 2 Vol., Handbücher zur Sprach- und Kommunikationswissenschaft 29, Berlin 2008f.
- Mahlberg, Michaela, English general nouns. A corpus theoretical approach, Studies in corpus linguistics 20, Amsterdam, Philadelphia 2005.
- **Mahlberg, Michaela, Corpus linguistics and discourse analysis, in: Schneider, Klaus P./Baaron, Anne (Hg.), Pragmatics of discourse, Handbook**

of Pragmatics 3, Berlin 2014, 215-238.

- Mukherjee, Joybrato, Anglistische Korpuslinguistik. Eine Einführung, Grundlagen der Anglistik und Amerikanistik 33, Berlin 2009.
- O'Keefe, Anne/McCarthy, Michael (Hg.), The Routledge handbook of corpus linguistics, Routledge handbooks in applied linguistics, London 2010.
- Perkuhn, Rainer/Keibel, Holger/Kupietz, Marc, Korpuslinguistik, UTB Sprachwissenschaft 3433, Paderborn 2012.
- Scherer, Carmen, Korpuslinguistik, Kurze Einführungen in die germanistische Linguistik 2, Heidelberg 2. Aufl. 2014.
- **Scott, Mike/Tribble, Christopher, Textual patterns. Key words and corpus analysis in language education, Studies in corpus linguistics 22, Amsterdam 2006.**
- Scott, Mike, WordSmith Tools Version 7, Liverpool 2016, <http://www.lexically.net/wordsmith/>; abgerufen am 08.10.2018.
- Sinclair, John, Trust the text. Language, corpus and discourse, London 2004.
- Sinclair, John, Corpus and text. Basic principles, in: Wynne, Martin (Hg.), Developing linguistic corpora. A guide to good practice, Oxford 2005, 1-16.
- Sinclair, John, Meaning in the framework of corpus linguistics [2005], in: Teubert, Wolfgang/Krishnamurthy, Ramesh (Hg.), Corpus linguistics. Vol. 1, Critical concepts in linguistics, London 2007, 182-196.
- Teubert, Wolfgang, Natural and human rights, work and property in the discourse of Catholic social doctrine, in: Hoey, Michael/Mahlberg, Michaela/Stubbs, Michael/Teubert, Wolfgang (Hg. u.a.), Text, discourse, and corpora. Theory and analysis, Studies in corpus and discourse, London 2007, 89-126.
- Tognini-Bonelli, Elena, Corpus linguistics at work, Studies in corpus linguistics 6, Amsterdam, Philadelphia 2001.
- Wynne, Martin, Art. Searching and concordancing, in: Corpus Linguistics. An international Handbook I (2008), 706-737.
- Ziebertz, Hans-Georg, Art. Empirische Religionspädagogik, in: Lexikon der Religionspädagogik II (2001), 1746-1750.

Abbildungsverzeichnis

- Abb. 1 Kollokationen des Schlüsselwortes Gott in WiReLex2018 © Stefan Altmeyer
- Abb. 2 Dimensionsmodell des sprachlichen Kontextes © Stefan Altmeyer
- Abb. 3 Ablaufschema einer sprachempirischen Theoriebildung © Stefan Altmeyer

Impressum

Hauptherausgeberinnen:

Prof. Dr. Mirjam Zimmermann (Universität Siegen)

Prof. Dr. Heike Lindner (Universität Köln)

„WiReLex“ ist ein Projekt der Deutschen Bibelgesellschaft

Deutsche Bibelgesellschaft

Balinger Straße 31 A

70567 Stuttgart

Deutschland

www.bibelwissenschaft.de