

Doris Höhmann*

Lexikographische Lösungsansätze: Zur Bedeutung korpuslinguistischer Kontextualisierungsstrategien

<https://doi.org/10.1515/glot-2017-0003>

Abstract: In recent years, both the availability of electronic resources (such as the BNC and the DWDS) and the possibility of compiling highly specialised large corpora (e.g. on the same topic, interaction type or communicative event) together with the development of ever more sophisticated corpus-linguistic tools, have opened new horizons in the study of language use and, consequently, in lexicography, translation and the teaching and learning of second/foreign languages. However, working with huge amounts of data and/or information remains a challenging issue, and search strategies are needed to cope more effectively with linguistic phenomena such as intralinguistic and interlinguistic variation in language use and polyfunctionality, which appear to be the rule rather than the exception in natural languages, not only in everyday language but in LSP communication, too.

This paper will focus on different contextualization strategies. The examples offered, drawn from highly specialized subcorpora on the one hand and from very large general language corpora on the other, will illustrate once more the impact of variables such as domain, topic, text type etc., which can be observed even in the case of core vocabulary items. Furthermore it will discuss the potential offered by concgrams, which allow us to cope with many of the difficulties described above.

Keywords: zwei- und mehrsprachige Lexikographie, Korpora, Mehrdeutigkeit/Ambiguität/Polysemie, Ko- und Kontext, Suchanfragen

1 Hintergrund

Zu den Grundproblemen der Lexikographie bzw. Lexikologie gehört die auf Nutzerbelange abgestimmte Aufbereitung lexikographischer Informationen. Wenngleich durch die technischen Fortschritte insgesamt deutliche Verbesserungen für die Nutzer und Nutzerinnen lexikographischer Nachschlagewerke

*Corresponding author: Doris Höhmann, Universität Sassari, Sassari, (Sardinien/Italien),
E-mail: dhoehmann@uniss.it

erzielt werden können (man denke etwa an die Möglichkeit, E-Wörterbücher mit bestehenden Datenbanken zu verlinken oder Korpuskomponenten einzubauen, terminologische Einträge durchgängig durch Bildmaterialien zu bereichern u.a.), bleibt das Verhältnis von gegebenem und gesuchtem Informationsangebot oft problematisch.

Auf der einen Seite bestehen gerade – aber nicht ausschließlich – im Bereich der fach- und domänenspezifischen Kommunikation nach wie vor Lücken in den erfassten Sprachdaten, wobei die Informationsdefizite nicht nur die Aufnahme bislang fehlender Lemmata oder weiterer Übersetzungsmöglichkeiten betreffen, sondern vor allem auch den weiten Bereich metasprachlicher Informationen zum Sprachgebrauch, etwa zur Wortkombinatorik, zur Abgrenzung zu Ausdrucksvarianten und dabei der Verwendung von Sprachmitteln verschiedener Sprachregister. Auf der anderen Seite werden die NutzerInnen mit einer Vielzahl an Informationen konfrontiert, aus der sie sich jeweils die gesuchten bzw. passenden Sprachmittel oder Bedeutungen herausfiltern müssen. Zwar können durch eine größere Menge an Informationen bestehende Informationsdefizite verringert werden, dabei ist aber das Risiko im Auge zu behalten, dass die Aufstockung des Informationsangebots nicht in Informationsüberfrachtung umschlägt.

Von dieser Dynamik ist auch die Möglichkeit betroffen, elektronische Korpora unterschiedlicher Größenordnung zusammenzustellen und als Komponente in lexikographische Online-Ressourcen einzubinden. Das Mehr an Sprachdaten, das durch den Einbau von Korpuskomponenten zur Verfügung steht, ist von kaum zu unterschätzendem Wert – zu den Vorteilen gehört insbesondere die größere Anzahl von Belegen, die die verschiedenen Lesarten bzw. Gebrauchsweisen eines Lexems veranschaulichen und Regelmäßigkeiten im Sprachgebrauch erst erkennen lassen – und gilt als unabdingbare Voraussetzung für eine Verbesserung lexikographischer Werke (u.a. Atkins/Rundell 2008). Doch stellt die Arbeit mit großen und sehr großen Korpora aufgrund der gegebenen Informationsfülle nach wie vor eine große Herausforderung dar. Die Versuche, eine Lösung für die hier auftretenden Probleme zu finden, sind vielfältiger Art. Zu nennen sind hier insbesondere die statistische Erfassung und Auswertung häufiger wiederkehrender Sprachmuster, etwa in Form von sogenannten Gebrauchs- und Verhaltensprofilen, *word sketches*, Wortprofilen, Kookkurrenzprofilen (u.a. Kilgarriff et al. 2004, Gries/Divjak 2009, Belica 2011, Geyken, 2011, Didakowski/Geyken 2013), sowie die sprachpflegerischen bzw. sprachdidaktischen Bemühungen, die Auswahl der angezeigten Konkordanzzeilen (KWICs¹) zunächst auf besonders typische Belege oder relativ einfach zu verstehende Beispiele einzugrenzen (u.a. Kilgarriff et al. 2008, Didakowski et al. 2012).

1 Key Words in Context.

In der lexikographischen Arbeit und in der empirischen sprachwissenschaftlichen Forschung, die sich ebenfalls mit dem Problem der Datenfülle konfrontiert sieht, wird die nur schwer zu bewältigende Menge an Sprachdaten zudem oft durch Stichproben, die Begrenzung der Kontexte auf eine bestimmte Größe (beispielsweise die gängige Beschränkung der Kollokationsspanne auf ≤ 5 Worteinheiten auf beiden Seiten des Suchworts) und/oder die Nichtberücksichtigung seltenerer Belege zurückgekürzt (u.a. Storrer 2006).

Die bei der Durchforstung umfangreicher Datenmengen auftretenden Schwierigkeiten scheinen dabei wesentlich an grundlegende Eigenschaften der natürlichen, historisch gewachsenen Kultursprachen gebunden zu sein. Es liegt daher nahe davon auszugehen, dass eine Optimierung der Recherchemöglichkeiten in erster Linie über eine bessere Kenntnis bzw. größere Berücksichtigung sprachlicher Organisations- und Funktionsprinzipien erreicht werden kann, die in ihrer Bedeutung zum Teil erst durch die – an den technischen Fortschritt der letzten Jahrzehnte geknüpften – korpuslinguistischen Untersuchungsmöglichkeiten deutlicher hervorgetreten sind. Dies gilt insbesondere für das Vorkommen zahlreicher Varianten, die anders als allgemein hin angenommen auch den Bereich der domänen- und fachspezifischen Kommunikation grundlegend prägen (dazu u.a. Cap/Heid 2011), und für den herausragenden Stellenwert der verschiedenen Formen der Ambiguität und Unterbestimmtheit auf Wortebene, die den Regelfall darzustellen scheinen.² Eine Verbesserung ist des Weiteren durch die computerlinguistische Umsetzung bereits vorliegender sprachwissenschaftlicher Erkenntnisse zu erwarten:

Firth's (1957: 6) suggestion to look at the "company that words keep" has been operationalised in a number of different ways (see Evert 2004, 2008, 2010) and has been explored in a number of different contexts (e.g. Baker et al. 2008, Xiao & McEnery 2006, Syanova & Schmitt 2008). However, more than fifty years into the research on collocations, many of the lessons learned from this research have yet to be systematically evaluated and fully implemented in the tools that corpus linguists use (Brezina et al. 2015, 139-140)

Vor diesem allgemeineren Hintergrund ist die Zielsetzung des vorliegenden Beitrags zu sehen, Selektionsmechanismen herauszuarbeiten bzw. weiterzuentwickeln, die es NutzerInnen von lexikographischen Ressourcen erlauben, „to get access, in the most intuitive way, to exactly the bit of information he/she needs in a given situation“ (Prinsloo u.a. 2012, 292). Um diesem Ziel näher zu kommen, soll im vorliegenden Beitrag der Stellenwert verschiedener Kontextualisierungsverfahren bzw. -strategien bei der Vereindeutigung und

2 Zur einzelwortbezogenen Ambiguität als semantischem Scheinproblem und zur im Weiteren verwendeten Begrifflichkeit, insbesondere zum Terminus ‚Polyfunktionalität‘, siehe Abschnitt 2.

sprachlich angemessenen Selektion polyfunktionaler Sprachbausteine oder Worteinheiten näher beleuchtet werden.

Zum besseren Verständnis soll vorab zunächst auf die Relevanz der Ko- und Kontextabhängigkeit von Bedeutungen bzw. Bedeutungsvarianten eingegangen werden.

2 Polyfunktionalität als Grundeigenschaft von Ausdrucksmitteln

Durch die Zusammenschau zahlreicher, aus ihrem weiteren Kommunikationszusammenhang herausgelösten Korpusbelegen (KWICs) wird die Bedeutung des Zusammenspiels von Ko- und Kontextkomponenten besonders augenfällig. Darüber hinaus wird deutlich, dass in konkreten Äußerungen das Problem der einzelwortbezogenen Ambiguität im Allgemeinen ein Scheinproblem darstellt, da polyseme und homonyme Ausdrücke (wie beispielsweise *Bank*, *Schloss*, *eben*, *bringen*, *geben*) zumeist schon auf der Grundlage des engeren Kontexts vereinheitlicht werden können (vgl. u.a. Stubbs 2002, Teubert 2005, Gross 2008).³

Durch die unterschiedliche Kombinierbarkeit von Sprachmitteln erhöhen sich – gemäß dem Prinzip der doppelten Artikulation – die Ausdrucksmöglichkeiten einer Sprache z.B. im Zuge begrifflicher Ausdifferenzierungen und Abgrenzungen auch auf der Ebene der Lexik. Zur näheren Erläuterung des Zusammenspiels von Ko- bzw. Kontextkomponenten sollen im Folgenden die Verbindungen herangezogen werden, die das Verb *spielen* mit weiteren Vokabeln aus dem Grundwortschatz eingeht.⁴

Die Beschränkung auf das isolierte Wort *spielen*, d.h. ohne Berücksichtigung seiner Einbettung in den sprachlichen Ko- und Kontext, würde insbesondere

³ Zu einer anderen Einschätzung gelangt Winkler (2009: 33), die in diesem Zusammenhang zwar ebenfalls von einer „grundlegende[n] Eigenschaft sprachlicher Ausdrücke“ spricht, diese jedoch darin sieht, dass sprachliche Ausdrücke „bei ihrer Verwendung in konkreten Äußerungen und Situationen oft nicht eindeutig interpretiert werden können“. Winkler rückt die auf diese Eigenschaft zurückgehenden, in ihrer Tragweite sicherlich nicht zu unterschätzenden Kommunikationsschwierigkeiten in den Vordergrund und nimmt damit eine andersgeartete Gewichtung des Problems der Vereindeutigung vor: Es sei die Vielzahl an Missverständnissen in der Kommunikation, die deutlich belege, dass Mehrdeutigkeit, Unterbestimmtheit, Vagheit oder Polysemie von sprachlichen Ausdrücken viel mehr der Normalfall sind als ihre Eindeutigkeit (vgl. ebenda).

⁴ Ausführlichere Darstellungen zu weiteren Beispielen finden sich u.a. in Stubbs 2002, Gries/Divjak 2009 und Höhmann 2011.

seinen stark polyfunktionalen Charakter verkennen. Wie die angeführten Beispiele zeigen und wie im Sprachvergleich durch den Rückgriff auf Übersetzungen in eine andere Sprache besonders deutlich wird, ist das Verb für sich allein genommen vom Prinzip her als semantisch unterdeterminiert einzustufen.⁵

- (1) *spielen*: typische Kotextkomponenten im Sprachvergleich (DE - IT)
- | | |
|----------------------------------|--|
| a. <i>Tennis spielen</i> | ital. <i>giocare a tennis</i> |
| b. <i>Gitarre spielen</i> | ital. <i>suonare la chitarra</i> |
| c. <i>die Hauptrolle spielen</i> | ital. <i>recitare la parte principale</i> ⁶ |
| d. <i>den Hamlet spielen</i> | ital. <i>interpretare, fare la parte di Amleto</i> |
| e. <i>eine Rolle spielen</i> | ital. <i>giocare un ruolo, svolgere un ruolo</i> |
- ...

Bei der Bedeutungseingrenzung bzw. Übersetzung von *spielen* in Verbindung mit Nomen wie *Gitarre*, *Klavier*, *Tennis*, *Ball*, *Schach* oder *Karten* ist bereits die Kenntnis der semantischen Klasse (Musikinstrument, Sportart/Spiel) für die Wahl des dazugehörigen verbalen Ausdrucks (*giocare* oder *suonare*) ausreichend. In den anderen Fällen liegen die Verhältnisse deutlich komplizierter, wenn etwa wie in den Beispielen (1) c. und (1) d. in den Einzelsprachen eine unterschiedliche begriffliche Differenzierung bei der Bestimmung der semantischen Klassen vorliegt. So würde der im Deutschen zur Charakterisierung ausreichende Oberbegriff „Theaterrolle“ zur Erfassung der Gebrauchsweisen der italienischen Verben *interpretare* und *recitare* nicht genügen.

Dabei stellt bekanntlich jede Ausgestaltung des Ko- bzw. Kontextes, einschließlich der Nicht-Realisierung von Äußerungskomponenten, eine solche Vereindeutigung der Verwendungsweisen dar.⁷ So bleibt im Fall der Verwendungsweise des Verbs ohne weitere Mitspieler und mit einer Person als AGENS

⁵ Zur Theoriebildung siehe u.a. die Ausführungen mit weiterführenden Literaturhinweisen in Winkler (2009), die sich jedoch im Wesentlichen auf die Bedeutungsvarianz von Lexemen durch metonymische Bedeutungsverschiebungen bezieht, und Roelcke (2004, 2010), der sich insbesondere mit der Ko- und Kontextabhängigkeit der Bedeutungen von Fachtermini auseinandersetzt.

⁶ Die beiden Beispiele c und d stammen aus der von der italienischen Tageszeitung *Corriere della Sera* veröffentlichten Online-Version des Wörterbuchs *Il Sansoni Tedesco. Deutsch-Italienisch, Italienisch-Deutsch* (http://dizionari.corriere.it/dizionario_tedesco/Tedesco/S/spielen.shtml)

⁷ Die Rolle, die die Bedeutung der syntaktischen Valenz bzw. syntaktischer Sprachmuster für die Bedeutungskonstituierung spielt, ist einer der Gründe dafür, weshalb eine strikte Trennung von Lexik, Semantik und Syntax vielfach als Sackgasse angesehen wird.

offen, welche Art des Spielens fokussiert wird.⁸ Die zum Verständnis des Vollverbs notwendige Information ist hier ebenfalls dem Ko- oder Kontext zu entnehmen:

(2) *x hat heute gespielt*

- a. Er hat heute gespielt, gekrabbelt, geplappert.

(Quelle: <http://www.urbia.de/archiv/forum/th-4745581/nach-b-menin-gokokken-impfung-hohes-fieber-bin-verzweifelt-vor-sorge.html>; Zugriff zuletzt am 1.9.2016)

- b. Daniel Zieker hat heute gespielt, als habe er nie gefehlt, da hatte ich während des Spieles regelrecht Gänsehaut.

(Quelle: http://www.tv-oppweiler.de/berichte/berichte_141011_1.htm; Zugriff zuletzt am 4.9.2016)⁹

Trotz zahlreicher Untersuchungen zu Bedeutungsvariationen, die auf Valenzunterschiede, metonymische Bedeutungsverschiebungen oder Bedeutungssetzungen zurückgehen, setzt sich die Erkenntnis, dass eine einzelwortzentrierte Auffassung von Lexik – wie auch eine strikte Trennung von Lexik und Grammatik – der Bedeutung des Zusammenspiels von Ko- und Kontextkomponenten nicht gerecht wird, erst langsam durch. Es scheint fast, als habe die in der Lexikographie lange Zeit vorherrschende, weit verbreitete einseitig einzelwortzentrierte Betrachtungsweise des allgemein- und fachsprachlichen Wortschatzes, der man im sprachlichen Alltag beispielsweise bei der Erstellung von Listen oder bei der Auflistung von Lemmata in Wörterbüchern vielfach begegnet, den Blick auf dieses Strukturprinzip verstellt. Dabei prägt das Zusammenspiel verschiedener Ko- und Kontextelemente die Bedeutungskonstituierung so wesentlich, dass ihm bei genauerer Betrachtung selbst Eigennamen und Zahlen, die als Inbegriff eindeutiger Referenz gelten,¹⁰ unterworfen sind:

8 Ein klassisches Beispiel für hier auftretende Bedeutungsunterschiede stellt das Verbs *trinken* dar: *x trinkt Kaffee* vs. *x trinkt*.

9 Auf dieses Beispiel wird zurückzukommen sein, veranschaulicht es doch, dass auch bei allgemeinsprachlichen Vokabeln der unmittelbare Kontext zur genaueren semantischen Eingrenzung (hier auf das Handball-Spielen) nicht immer ausreichend ist. Das zur weiteren Klärung des Verbs *spielen* (als *Amateurhandballer* oder *Profihandballer*) notwendige Wissen kann dem Text selbst nicht entnommen bzw. nur bei ausreichendem Hintergrundwissen über Daniel Zieker und die genannten Vereine (TV Oppweiler, HG Oftersheim/Schwetzingen) aktiviert werden.

10 Neben der Kontextualisierung durch Maßeinheiten und andere Bezugsgrößen ist auch der kulturspezifische Symbolwert zu nennen, den Zahlen besitzen (s. dazu u.a. Keßler 2008).

- (3) *Martin – Martin Luther – Martin Luther King*
- (4) *Lüneburger Heide – Erlebniswelt Lüneburger Heide GmbH*
- (5) *10 Meter – 10 Millimeter – 10⁻⁶ Millimeter*

Um dem Umstand Rechnung zu tragen, dass Worteinheiten im Normalfall je nach eingegangener Verbindung verschiedene Funktionen erfüllen können und als Komponenten an der Konstituierung unterschiedlicher Bedeutungen beteiligt sind,¹¹ und schließlich um Assoziationen zu vermeiden, die dieses Strukturprinzip in erster Linie mit dem Makel mangelnder Klarheit oder einer absichtlich unklaren oder doppeldeutigen Ausdrucksweise in Verbindung bringen, scheint es dabei von Vorteil zu sein, an Stelle der Termini ‚Mehrdeutigkeit‘ und ‚Ambiguität‘ den neutralen Terminus Polyfunktionalität als Oberbegriff zu setzen. Durch seine Verwendung wird zudem das Problem der Klassifizierung unterschiedlicher Formen der Unterbestimmtheit (Polysemie, Homonymie und andere Formen der wortbezogenen Ambiguität) umgangen, deren genaue Unterscheidung oft nicht mit Bestimmtheit vorgenommen werden kann und für synchrone Sprachuntersuchungen sowie für die konkrete Vereindeutigung (etwa bei der Übersetzung in eine andere Sprache oder im Rahmen der Durchforstung umfangreicherer Datenmengen) im Allgemeinen zweitrangig ist.¹² Die leicht zu vereindeutigenden bzw. zu kontextualisierenden lexikalischen Komponenten sollten dabei deutlich von vagen und ambig bleibenden, nicht oder nur schwer zu klärenden Ausdrucksweisen, die zu Missverständnissen führen können, unterschieden werden.

Die mit der Polyfunktionalität einhergehende Polylexikalität von Ausdrucksmitteln stellt dabei gerade bei einer Klammersprache wie dem Deutschen eine besondere Herausforderung dar, wie anhand des Konstrukts *Rolle + spielen* aufgezeigt werden soll. Die folgenden Beispielbelege (6) für umfangreichere Satzklammern, die verschiedenen Themenbereichen und/oder Textsorten entnommen wurden, lassen dabei zum einen deutlich werden, dass eine Kollokationsspanne von ≤ 5 Worteinheiten in vielen Fällen der Syntax des Deutschen als Klammersprache nicht gerecht werden kann. Zum anderen veranschaulichen sie die herausragende Bedeutung der Begriffe Vor- und Nachverb (Weinrich

¹¹ Aufschlussreich ist in diesem Zusammenhang bereits ein bewusster Blick in ein größeres zweisprachiges Wörterbuch.

¹² Zur Bedeutung der Unterscheidung von Polysemie und Homonymie und ihrer Problematik siehe u.a. Lyons (1968, 406–407) und Gries (2015).

2003) für die Untersuchung komplexer Äußerungen, die weit über die übliche Unterscheidung trennbarer und untrennbarer Verben hinausführt.

- (6) *Rolle + spielen*: Beispiele für umfangreichere Satzklammern
- a. Der ‚rationale Kern‘ des nationalsozialistischen Antijudaismus bestand in der Faktizität der großen **Rolle**, die eine Anzahl von einzelnen Persönlichkeiten jüdischer Herkunft – offensichtlich nicht zuletzt wegen der universalistischen und messianischen Traditionen des Judentums – in der sozialistischen und kommunistischen Weltbewegung **spielte**.
(Quelle: Die Zeit, 26.02.1998, Nr. 10; DWDS-Korpus)
 - b. Welche **Rolle** bei dieser gegen die Sowjetunion und zeitweise auch gegen China gerichteten Vernichtungsstrategie Hitlers alte Garde **spielte**, hat Christopher Simpson 1988 dokumentiert
(Quelle: Hannover, Heinrich. 2001 [1998]. Die Republik vor Gericht 1954 - 1974, Berlin: Aufbau-Taschenbuch-Verl., 73; DWDS-Korpus)
 - c. Bei der anschließenden Diskussion in der Feintrassierung **spielte** zugunsten der Nordvarianten im Verhältnis zu den Südvarianten zwar auch die angenommene größere Entlastung des Ortskerns O. vom Durchgangsverkehr eine **Rolle**. (Quelle: OVG Münster Urteil vom 13.04.2011 - 11 D 37/10.AK; <https://openjur.de/u/448311.html>)
 - d. Allerdings **spielt** die Verarbeitung von Kumulations- oder Verlagerungseffekten sowie von Wechselwirkungen zwischen verschiedenen Umweltmedien auf der Ebene der SUP eine wesentlich größere **Rolle** (Meßerschmidt, Europäisches Umweltrecht, 2011, § 8, Rdnr. 135).
(Quelle: Gärditz, Klaus Ferdinand. 2016. Vertiefung Baurecht, Raumordnungs- und Fachplanungsrecht. Sommersemester 2016; https://www.jura.uni-bonn.de/fileadmin/Fachbereich_Rechtswissenschaft/Einrichtungen/Lehrstuehle/Gaerditz/Vorlesung/PlanungsR/BauRII-AP2.pdf; Zugriff zuletzt am 3.9.2016)

Festzuhalten ist, dass Polyfunktionalität¹³ und Polylexikalität als korpuslinguistische Probleme zwar gelöst werden können, sie jedoch die Durchforstung bzw. Untersuchung umfangreicherer Sprachbestände mitunter sehr erschweren. Sie

¹³ Gross (2008) sieht sich aufgrund der vereindeutigenden Wirkung des Zusammenspiels von Sprachmitteln zur Auffassung veranlasst, dass Polysemie im Grunde genommen eine Erfindung der Sprachwissenschaft sei.

verlangsamen sie nicht nur, sondern machen sie bei semantisch bzw. begrifflich und syntaktisch komplexeren Äußerungen auch fehleranfällig.¹⁴

In puncto Vereindeutigung bzw. Kontextualisierung sind Abstriche gerade in Hinblick auf die korpuslinguistische Untersuchung fachsprachlicher Texte zu machen. Man denke etwa an die zahlreichen Begriffsabgrenzungen und Bedeutungssetzungen, die beispielsweise im philosophischen Diskurs oder in der juristischen Kommunikation¹⁵ mit ein und derselben Bezeichnung (selbst innerhalb eines einzigen Textes) verknüpft sein können. Inhaltlich oft nur schwer voneinander abzugrenzen sind beispielsweise die verschiedenen begrifflichen Varianten des Lexems *Anschauung* (vgl. Roelcke 1989), die darüber hinaus erst einmal als Fachtermini erkannt werden müssen. Zu nennen ist aber auch das Vorkommen fachsprachlicher Termini in Textbeständen, die sich vorwiegend an ein Publikum aus interessierten Laien richten. Der folgende Beleg aus einem mehrsprachigen Korpus zur Tourismuskommunikation¹⁶ veranschaulicht beispielsweise die bei der Bedeutungsbestimmung des Lexems *National Park* auftretenden begrifflichen Komplikationen:

- (7) The IUCN category II is called ‘National Parks’. But the UKs National Parks are not in this category, because they are places where people live and work. (Quelle: <http://www.nationalparks.gov.uk/students/whatisanationalpark/nationalparksareprotectedareas/nationalparksaroundtheworld>; Zugriff zuletzt am 4.9.2016)

Auch die Abkürzung in dem Beispielbeleg (*IUCN – International Union for Conservation of Nature*) will erst einmal aufgelöst sein. Sie findet sich zwar auf derselben Webseite, doch mehrere Sätze bzw. Absätze (insgesamt 173 Worttokens) weit entfernt. Damit ist das zweite Problem angesprochen, das den Vereindeutigungsprozess auch in nicht fachsprachlichen Textbeständen sehr erschweren kann.

Bei der Durchforstung großer Textsammlungen bzw. der Zusammenschau zahlreicher Beispielbelege aus verschiedenen Quellen wird die Organisationsform ‚Text‘ aufgehoben, die wesentlich zur Bedeutungseingrenzung beiträgt und im Allgemeinen über den näheren (wenn auch nicht immer unmittelbaren) Ko- und Kontext eine Vereindeutigung erlauben würde. Wesentliche Komponenten sind

¹⁴ Siehe dazu u.a. Faaß et al. 2009 und Kübler 2010.

¹⁵ Zu den besonderen Schwierigkeiten quantitativ-qualitativ ausgerichteter Untersuchungen der Rechtssprache s. die einführende Darstellung in Höhmann 2011.

¹⁶ Die Zusammenstellung des Korpus wird in Fußnote 42 näher erläutert.

hierbei insbesondere Überschriften, die Untergliederung in Textabschnitte sowie das gegebenenfalls beigefügte Bildmaterial.

Die Schwierigkeiten der Bedeutungsbestimmung, die durch die Art der Datenaufbereitung bzw. bei Nichtbeachtung der texteigenen Ordnungsstruktur auftreten, soll anhand des Beispiel (2b) [*Daniel Zieker hat heute gespielt, als habe er nie gefehlt, da hatte ich während des Spieles regelrecht Gänsehaut.*] veranschaulicht werden, das hier noch einmal in einen umfangreicheren Kontext eingebettet wiedergegeben wird (8).

(8) *spielen* – Beispielbeleg zur Einbettung in den Textzusammenhang

Oppenweilers **Handballer** weiter im Höhenflug: Deutlicher Heimsieg gegen Mitaufstiegs kandidaten

[...]

Als Kevin Matschke das 33:21 erzielte, hatten sich alle TVO-**Handballer** in die Torschützenliste eingetragen.

[182 Tokens ohne eine für Handball-Laien ausreichende Vereindeutigung durch Ko- bzw. Kontextkomponenten]

Volker Blumenschein: »Die **HG** Oftersheim/Schwetzingen spielt schon eine sehr außergewöhnliche Deckung. Das ist schwer zu bespielen, und deshalb haben wir anfangs auch ein bisschen gebraucht. Dann war unser Abschlussverhalten sehr gut und wir haben oft bis an den Kreis gespielt. In der zweiten Halbzeit hatten wir 19 Tore bei 33 Angriffen, das ist ein hervorragender Wert. Daniel Zieker hat heute **gespielt**, als habe er nie gefehlt, da hatte ich während des Spieles regelrecht Gänsehaut. Und bei Stefan Merzbacher ist es wie bei einem 21-Jährigen, der will einfach unbedingt spielen. So hatten wir heute eine tolle Torhüterleistung: Erst hat Tobias Barthold überragend gehalten, dann kam Thomas Fink und hat überragend gehalten, und dann hat Stefan Merzbacher auch noch einige Bälle abgewehrt.«

[weitere 71 für eine Vereindeutigung nicht ausreichende Tokens]

(Quelle: http://www.tv-oppenweiler.de/berichte/berichte_141011_1.htm; Zugriff zuletzt am 1.9.2016)

Für die weitere Vereindeutigung des Verbs *spielen* würden die Schlagzeile sowie das eingefügte Foto des Spielers ausreichen. Ist man hingegen gezwungen, sich aufgrund der Datenaufbereitung allein auf die nähere sprachliche Umgebung zu stützen, sind für Handball-Laien bis zum nächsten Vorkommen des Lexems *Handball* eine Kollokationsspanne von 246 Tokens zu überbrücken, für Sportfans, die die Abkürzung *HG* (Handball-Gemeinschaft) kennen oder denen die

„HG Oftersheim/Schwetzingen“ ein Begriff ist, fällt der Abstand geringer aus, umfasst aber immerhin noch 56 Worttokens bzw. 3 ganze Sätze. Die weiteren semantisch eingrenzend wirkenden Vokabeln (wie *Halbzeit*, *Tore*, *Bälle*, *Torhüterleistung*) sind allein für sich genommen nicht ausreichend, da sie auch im Zusammenhang mit anderen (Ball-)Sportarten verwendet werden.¹⁷

Mit der Berücksichtigung (oder Nicht-Berücksichtigung) der Organisationsform ‚Text‘ bei der Aufbereitung von Sprachdaten ist das Problem der Kontextualisierungsstrategien angeschnitten, auf die sich das Hauptaugenmerk der vorliegenden Studie richtet. Im Folgenden wird auf die Kontextualisierung von Sprachmitteln durch die Variablen Textsorte und Themengebiet sowie auf die eingrenzende Wirkung innersprachlicher Organisationsprinzipien näher eingegangen. Dabei wird angesichts des Umstands, dass Suchanfragen in den größeren Zeitungs- und Webkorpora zu mehr Treffern führen können als in den weitaus weniger umfangreichen, dafür aber unvergleichlich zeit- und kostenintensiver aufgebauten Referenzkorpora (s. dazu u.a. Baroni 2010), die Frage zu beantworten sein, wie zeitgemäß bzw. zukunftssträftig die Zusammenstellung hochspezialisierter Korpora noch ist.

3 (Teil-)Fachgebiet

Die Vorzüge hochspezialisierter Korpora für terminographische Erhebungen, aber auch für diskursanalytische Untersuchungen dürften kaum zur Debatte stehen. Doch auch sonst, etwa beim Nachschlagen von Vokabeln oder beim Auffinden einzelner Belegstellen überwiegen die Vorteile eines domänen- bzw. fachspezifischen Datenzuschnitts (in Form eines sorgfältigen Korpusdesigns oder auf der Grundlage einer gezielten Auswahl aus umfangreicheren Textbeständen), wie die Ergebnisse der Evaluierung eines kleineren, dafür aber hochspezialisierten Korpus zeigen. In einer dazu durchgeführten Studie wurde das „Renaissance-Korpus“, ein insgesamt ca. 1 Million Tokens umfassendes dreisprachiges Korpus aus Texten zur darstellenden Kunst in der Renaissance,¹⁸ in Hinblick auf seine Ergiebigkeit mit den großen online zugänglichen Korpora des

¹⁷ Ein weiteres ausführliches Beispiel („Biennale“) findet sich in Höhmann (2014a).

¹⁸ Das Korpus wurde in Zusammenarbeit mit den Kunsthistorikerinnen Caterina Virdis und Maria Vittoria Spissu im Rahmen eines von der Regione Sardegna geförderten Forschungsprojekts zur quantitativ-qualitativen Untersuchung authentischer Textvorkommen (2010-2012) erstellt. Eine ausführlichere Beschreibung des Korpus und seiner Verwendungsmöglichkeiten zu wissenschaftlichen und didaktischen Zwecken nicht nur in der sprachwissenschaftlichen Forschung und Fremdsprachendidaktik, sondern auch in der Kunstwissenschaft findet sich in Höhmann/Spissu (2013).

Deutschen und Italienischen verglichen, deren Größenordnungen im dreistelligen Millionenbereich und höher liegen (Stand 2013/2014).¹⁹ Untersucht wurde im Einzelnen das Vorkommen von deutschen und italienischen Lexemen aus dem Lexemverband „Donauschule“.²⁰ Dabei erzielte das kunstgeschichtliche Korpus trotz seines deutlich kleineren Umfangs mit insgesamt 46:26 lexikalischen Einheiten nicht nur in Bezug auf die erfassten Ausdrucksmittel entschieden bessere Werte als die herangezogenen Onlineressourcen, sondern schnitt vor allem auch hinsichtlich des Verhältnisses von relevanten und nicht relevanten Treffern (wahren und falschen Positiven) erheblich besser ab, und damit auch im Hinblick auf den Arbeitsaufwand, den die manuelle Durchsicht der Gesamtmenge aller Treffer erfordert. Während in den beiden deutschen und italienischen Teilkorpora die Quote der relevanten Treffer bei jeweils ca. 85% lag, überschritt sie bei den Großkorpora im günstigen Fall knapp die 1%-Marke. Die Aussortierung nicht relevanter Treffer wurde dabei auch hier durch polyfunktionale lexikalische Einheiten mit zwei oder mehr Lesarten erschwert. Dass bei der nicht weiter gefilterten Zusammenschau von Konkordanzzeilen (KWICs) die Wissensframes aufgebrochen werden, innerhalb derer die Korpusbelege in den ursprünglichen kommunikativen Zusammenhängen verortet und kontextualisiert sind, äußerte sich hier hauptsächlich darin, dass von den erfassten lexikalischen Einheiten unterschiedliche Domänen berührt werden: Genannt seien beispielsweise das Adjektiv *donauländisch*, das u.a. zur Kennzeichnung von Artefakten aus früheren, prähistorischen Epochen Verwendung findet, und das Substantiv *Donaulandschaft*, das sowohl eine natürliche Landschaft, eine bildliche Landschaftsdarstellung und schließlich das Genre bezeichnen kann, wobei als weitere Verkomplizierung die Bilder auch später entstanden sein können und daher nicht unbedingt der Donauschule zuzuordnen sein müssen. Zudem stimmen die aktualisierten Bedeutungen in den beiden Sprachen nicht

19 Folgende Korpora wurden miteinander verglichen: die für beide Sprachen erstellten TenTen-Korpora (Jakubíček et al. 2013; <http://www.sketchengine.co.uk>), für das Deutsche des Weiteren die Referenz- und Zeitungskorpora, die am Institut für Deutsche Sprache in Mannheim (Kupietz et al. 2010; <https://cosmas2.ids-mannheim.de/cosmas2-web/>) und an der Berlin-Brandenburgischen Akademie der Wissenschaften (Geyken 2011; <http://www.dwds.de/>) beheimatet sind, für das Italienische die Risorsa Dinamica Italiana di Rete RIDIRE, ein nach Gegenstandsbereichen untergliedertes Webkorpus (Panunzi et al. 2014; <http://www.ridire.it/it.drwolf.ridire/home.seam>), das Referenzkorpus CORpus di Italiano Scritto CORIS (Rossini Favretti 2000; http://corpora.dslo.unibo.it/coris_ita.html) sowie das Zeitungskorpus Corpus della Repubblica (Baroni et al. 2004; <http://sslmit.unibo.it/repubblica/>).

20 Detailliertere Untersuchungsergebnisse finden sich in der Studie “Il “corpus del rinascimento”. Stato dell’arte e prospettive” (Höhmann 2016).

immer überein: *Donauschule* kommt in den deutschen Belegen auch als Bezeichnung von Bildungseinrichtungen vor, in den italienischen Belegen (*Scuola del Danubio/danubiana*) auch für einen besonderen Stil des Fußballspiels. Eine besondere Komplikation stellt dabei dar, dass in den polyfunktionalen Konstrukten (*stile danubiano* oder *maestri della scuola danubiana*) lexikalische Elemente (*stile*, *maestro*, *scuola*) enthalten sind, die im kunsthistorischen Diskurs verwendet werden und bei der Eingrenzung kunstgeschichtlicher Textbestände zum Einsatz kommen können.

Einschränkend zu den Befunden ist anzumerken, dass die Zusammenstellung von Korpora bekanntlich wesentlich von Fragen des Urheberrechts, aber auch praktischen Problemen wie der bereits geleisteten oder noch vorzunehmenden Digitalisierung von Texten mitbestimmt wird (s. dazu u.a. Kupietz/Lüngen 2014). Je nachdem, welche domänen- bzw. fachbezogenen Sprachbestände in die Ressourcen aufgenommen bzw. nicht aufgenommen werden konnten, kann das Ergebnis einer solchen Evaluierung entsprechend unterschiedlich ausfallen.

4 Textart

Die Bedeutung der Variable Textart, deren stärkere Berücksichtigung insbesondere in der Fachlexikographie zu den noch kaum umgesetzten Desiderata gehört, spielt ebenfalls eine nicht zu unterschätzende Rolle bei der lexikographischen Aufbereitung von Sprachbeständen. Als „eingefrorene“ Konstellation von Faktoren der Kommunikationssituation erlaubt es die Textartenzugehörigkeit in Verbindung mit dem Gegenstandsbereich, kommunikative Handlungsräume weiter einzuschränken und die darin vorkommenden Sprachbestände unter einer weitergehenden Wahrung der kommunikativen Zusammenhänge einzugrenzen, als dies sonst der Fall ist. Bei der Untersuchung einer entsprechend großen Datengrundlage zeichnen sich deutlich textartenbezogene Tendenzen im Sprachgebrauch ab, in denen sich die in den verschiedenen Textarten jeweils erfüllten kommunikativen Aufgaben niederschlagen. Die Unterschiede betreffen dabei nicht nur die Erfassung von Fachtermini im engeren Sinne, sondern auch die Schnittmenge aus allgemein- und fachsprachlichen Sprachbeständen.

Ein geradezu klassisches Beispiel, das es ermöglicht, „erhebliche textsortenspezifische Gebrauchsfrequenz-Unterschiede“ (Brandt 1996, 237) aufzuzeigen, ist die Untersuchung der Vorkommenshäufigkeit von Modalverben und anderen Modalitätsmarkern in Gesetzestexten, die hier aufgegriffen und auf die Textarten ‚Gerichtsentscheidung‘ und ‚Gesetzeskommentar‘ ausgeweitet wird. Zu diesem Zweck wird ein ca. 1,3 Mio. Tokens großes in sich geschichtetes,

hochspezialisiertes Korpus des deutschen und italienischen Verwaltungsrechts hinzugezogen.²¹

Bereits die auf den Kernbereich der Modalverben *können*, *müssen*, *sollen*, *wollen* und *dürfen* beschränkte Erfassung der Gebrauchsfrequenzen erlaubt die Feststellung textartenspezifischer Unterschiede (siehe Tabelle 1). So sticht etwa das deutlich höhere Vorkommen von *dürfen* in den verwaltungsrechtlichen Textbeständen und dabei insbesondere in den Textarten Gesetzestext und Gesetzeskommentar ins Auge. Die Begrenzung der Daten gestattet zugleich einen direkteren Vergleich mit den Untersuchungsergebnissen Brandts (1996, 237). Die von ihm auf der Grundlage von 16 in Auszügen oder vollständig erfassten Gesetzestexten aus der zweiten Hälfte des 20. Jahrhunderts ermittelten Gebrauchsfrequenzen (*können/möchten*: 60,6%; *müssen*: 8,9%; *sollen*: 14,2%; *dürfen*: 14,0%; *wollen*: 1,3%) weichen von den hier erzielten Werten zum Teil

Tabelle 1: Textartenspezifische Verteilung einiger Modalverben im innerdeutschen Sprachvergleich.

	DWDS-Zeitungskorpora ²²	Korpus des deutschen Verwaltungsrechts ²³			
		Gesamtkorpus	TK 1	TK 2	TK 3
<i>können</i>	35,0 %	56,6 %	63,1%	55,0 %	49,1 %
<i>müssen</i>	21,3 %	16,2 %	9,8 %	19,1 %	22,5 %
<i>sollen</i>	19,8 %	9,4 %	9,2 %	11,0 %	8,2 %
<i>wollen</i>	18,0 %	2,4 %	0,7 %	4,0 %	3,2 %
<i>dürfen</i>	5,9 %	15,4 %	17,2 %	10,9 %	17,0 %

²¹ Bei der Zusammenstellung der Subkorpora wurden möglichst viele Variablen der Kommunikationssituation konstant gehalten, so dass das Korpusdesign es erlaubt, die Auswirkung der Variable Textart auf die Beschaffenheit von Sprachbeständen genauer zu untersuchen. Neben den Kriterien Rechtssprache und Textart – die Subkorpora umfassen für die deutsche und italienische Rechtssprache Gesetzestexte und Gerichtsentscheidungen der höchsten Instanz sowie zusätzlich für das deutsche Teilkorpus Gesetzeskommentare – wurden insbesondere Fach- bzw. Teilfachgebiet und engerer Gegenstandsbereich („verwaltungsrechtliches Ermessen“) und Fachlichkeitsgrad berücksichtigt, um auf diese Weise eine hohe kommunikative Dichte zu erreichen. Zur letzteren trägt bei, dass die drei ausgewählten Textarten im beruflichen Alltag von JuristInnen gemeinsam verwendet werden und eine abgestufte Normativität aufweisen (zur Unterscheidung verschiedener „Grade und Ausformungen der Normativität“ in juristischen Textarten vgl. Busse 2000: 661). Genauere Angaben zu den Textbeständen finden sich in Höhmann (2011).

Die kommunikative Dichte von Korpusbeständen, die sich im vorliegenden Fall u.a. in zahlreichen intertextuellen Bezügen äußert, dient der systematischeren Erfassung der Ausdrucksmittel, die zur Bewältigung bestimmter kommunikativer Aufgaben dienen.

²² Stand September 2016 (www.dwds.de).

²³ S. Fußnote 21; TK 1 = Gesetzestexte, TK 2 = Gerichtsentscheidungen, TK 3 = Gesetzeskommentare.

deutlich ab, und zwar sowohl in Hinblick auf die Rangfolge der Modalverben als auch hinsichtlich ihrer Häufigkeit. Die feststellbaren Abweichungen können zwar als mögliches Indiz für die Auswirkung der Variable „Teilfachgebiet Verwaltungsrecht“ gesehen werden, lassen es jedoch vor allem wünschenswert erscheinen, auch hier mit umfangreicheren Datengrundlagen arbeiten zu können.

Dass die textartenspezifischen Unterschiede gerade auch für die zweisprachige Lexikographie von Bedeutung sind, zeigen beispielsweise die unterschiedlichen Gebrauchsfrequenzen des modalen Infinitivs und der italienischen Infinitivkonstruktion ‘*da + infinito*’ (siehe Tabelle 2).²⁴ Dabei werden im zwischen sprachlichen Vergleich sowohl einzelsprachliche als auch textartenspezifische Merkmale bzw. Gebrauchstendenzen sichtbar: Anders als in den deutschen Subkorpora, in denen der modale Infinitiv von der Rangfolge her noch vor *können* liegt,²⁵ fällt in den italienischen Textbeständen die Gebrauchsfrequenz der Infinitivkonstruktion nicht nur insgesamt viel niedriger aus, sondern die in den deutschen Gesetzestexten besonders ausgeprägte Verwendungsweise mit finitem Kopulaverb fehlt in den italienischen fast ganz und ist auch in den italienischen Gesetzestexten vergleichsweise selten belegt. In sprachvergleichender Hinsicht ist schließlich interessant, dass in den deutschen Texten die Gebrauchsvariante des

Tabelle 2: Modaler Infinitiv / ‘*da + infinito*’ im inner- und zwischensprachlichen Vergleich.²⁶

	italienische Gesetzestexte	Entscheidungen des CdS*	deutsche Gesetzestexte	Entscheidungen des BVerwG**
insgesamt	388	349	2868	1246
davon				
mit finiter Verbform	2	45	2626	1012
ohne finite Verbform	387	304	242	234

*Consiglio di Stato **Bundesverwaltungsgericht

²⁴ Das Beispiel (einschließlich der Tabelle) wurde einer früheren Arbeit der Verfasserin entnommen (Höhmann 2007). Aufgrund der bestehenden Umformungsmöglichkeiten (*die zu genehmigende Maßnahme* → *die Maßnahme, die zu genehmigen ist*) und besseren zwischensprachlichen Vergleichsmöglichkeiten wird das Modalpartizip im Anschluss an Holl (2001) als Variante des modalen Infinitivs aufgefasst.

²⁵ Zum Vergleich: *können* kommt in den beiden Subkorpora Gesetzestexte und Gerichtsentscheidungen 1934- bzw. 1114-mal vor. Zur hohen Gebrauchsfrequenz des modalen Infinitivs in Gesetzestexten s. u. a. Brandt (1996) und Soffritti (2000).

²⁶ Alle Subkorpora haben den ungefähr gleichen Umfang von 250000-300000 Tokens, so dass die Werte direkt miteinander vergleichbar sind.

modalen Infinitivs mit expliziertem AGENS (*'haben + zu + Infinitiv'*) ca. 18% der insgesamt 4114 Belege ausmacht, während in den italienischen Texten die entsprechende Variante nicht vorkommt.

So wichtig die sich hier abzeichnenden Gebrauchstendenzen für die Eingrenzung von in bestimmten Kommunikationszusammenhängen vorkommenden Sprachbeständen und ihre genauere Untersuchung sind, erweisen sich Angaben dieser Art nicht als konkret genug, um im Einzelfall a) die für eine bestimmte kommunikative Aufgabe gebrauchten Ausdrucksmittel zu ermitteln und b) die kommunikativen Aufgaben genauer zu verstehen, für die bestimmte Sprachmittel verwendet werden. Eine genauere kommunikative Bestimmung der Gebrauchsweisen erweist sich gerade im Fall von Modalverben und ihren Ersatzformen angesichts ihrer komplexen Bedeutungskonstituierung als ausgesprochen schwierig. Zu denken ist hierbei nicht nur an die herausragende Rolle, die die Negation bei Modalitätsmarkern spielt (z.B. *dürfen* vs. *nicht dürfen*), an den oft nur schwer zu bestimmenden Verbindlichkeitsgrad von *sollen* in direktiven Sprechhandlungen oder an die nur schwer auflösbaren oder bestehenbleibenden Ambiguitäten, die beim Gebrauch des modalen Infinitivs auftreten, der bekanntlich je nach Zusammenhang eine Notwendigkeit oder eine Möglichkeit oder beides zum Ausdruck bringen kann (z.B. *es ist zu erwarten*). Auch sonst lässt sich eine Vielzahl verschiedener Gebrauchsmuster bzw. Funktionen (u.a. epistemisch vs. deontisch) unterscheiden.

Wesentliche Verbesserungen lassen sich hier, wie im Folgenden ausgeführt werden soll, über die Wirkung sprachlicher Organisationsprinzipien erzielen, die noch unterhalb der Textebene operieren. Das Desiderat, Recherchen auf der Grundlage möglichst umfangreicher und dabei einschlägiger Sprachbestände durchzuführen, bleibt davon unberührt: Wie wünschenswert eine sowohl in quantitativer als auch in qualitativer Hinsicht zufriedenstellende Datengrundlage ist, tritt beim Einsatz weiterer Kontextualisierungsverfahren vielmehr noch deutlicher hervor.

5 Komponenten der sprachlichen Umgebung: ko(n)textuell miteinander verbundene Wörter bzw. Worteinheiten

Die Einsicht, dass viele der praktischen Schwierigkeiten der Korpusrecherche an das Zusammenspiel verschiedener, tendenziell polyfunktionaler und dabei oft polylexikalischer Äußerungskomponenten gebunden sind, legt es nahe, die verdeutlichende bzw. kontextualisierende Wirkung der unmittelbareren sprachlichen Umgebung näher zu untersuchen und, wo möglich, zu operationalisieren. Durch diese Herangehensweise wird es möglich, Kontextualisierungsstrategien zu entwickeln bzw. weiter zu entwickeln, denen neben der linearen Abfolge auch andere sprachliche Strukturierungsprinzipien zugrunde liegen.

Die Frage nach der vereindeutigenden bzw. kontextualisierenden Wirkungsweise des näheren sprachlichen Umfeldes führt im Wesentlichen zu einer Fokusverschiebung bei der Erfassung von Sprachmitteln: Gefragt wird nach den Komponenten des sprachlichen Umfelds, zu denen eine engere oder besonders enge, vereindeutigend wirkende Verbindung zu verzeichnen ist. Welche Art der Verbesserung auf diese Weise bei der Verknüpfung quantitativer und qualitativer Untersuchungsmethoden erzielt werden kann, sei zunächst anhand grammatikalischer Bindungen²⁷ veranschaulicht.

5.1 Grammatikalische Bindungen

Um das obige Beispiel weiterzuführen, besteht im Fall von Modalverben bzw. Modalitätsmarkern grammatisch gesehen eine der engsten Verbindungen zu dem von ihm abhängigen Infinitiv. Die Wahl des verbalen Lexems im Infinitiv als Ausgangspunkt hat dabei mehrere Vorteile: Mikroskopisch feine Regelmäßigkeiten im Sprachgebrauch werden sichtbar und können zum Teil schon bei einer ersten Durchsicht erfasst werden. So fallen beim Modalisierungsverhalten der Verben *begründen* und *erteilen* (s. Tabelle 3) beispielsweise unterschiedliche Gebrauchstendenzen bei *dürfen* und *können* auf: Während die beiden Modalverben deutlich häufiger zusammen mit *erteilen* vorkommen und dabei am häufigsten in den Gesetzestexten anzutreffen sind, wird *begründen* insgesamt nur zweimal mit *dürfen* und mit *können* hauptsächlich in den untersuchten Gerichtsentscheidungen und Gesetzeskommentaren zusammen verwendet. Bei *erteilen* ist im Vergleich zu *begründen* auch der Gebrauch weiterer Modallexeme (wie z.B. *verlangen* oder *geeignet sein*) stärker ausgeprägt, wobei sich auch hier wieder textartenspezifische Unterschiede feststellen lassen.

Ein wesentlicher methodischer Vorzug dieser Vorgehensweise besteht darüber hinaus darin, dass die gesamte Bandbreite der vorkommenden potenziellen Modalitätsmarker in ihrer Vielfalt aufscheint, d.h. keine Vorabauswahl getroffen wird und Klassifizierungsfragen zunächst zurückgestellt werden können. Zudem erleichtert diese Art des Datenzuschnitts, in weiteren Arbeitsschritten eine genauere Untersuchung der einzelnen Treffer vorzunehmen: Insbesondere können

²⁷ Stark vereindeutigend wirkt bereits die Wortartenzugehörigkeit (z.B. *leben* vs. *Leben*), die jedoch im Vergleich zum Englischen – der Sprache, für die in der Anfangsphase der Korpuslinguistik die meisten Tools entwickelt wurden – im Deutschen eine weitaus geringere Rolle spielt. Zur Notwendigkeit, bei der Entwicklung korpuslinguistischer Werkzeuge auf einzelsprachliche Strukturmerkmale einzugehen, siehe auch weiter unten.

implizite Modalisierungen²⁸ leichter erfasst werden, wodurch auch das Verhältnis von expliziter und impliziter Modalisierung der Untersuchung zugänglich wird, und schließlich können die aktualisierten Gebrauchsweisen genauer bestimmt werden. Letzteres ist angesichts der komplexen Bedeutungskonstituierung bei Modalverben und ihren Ersatzformen besonders relevant.

Tabelle 3: Eingegangene Verknüpfungen: das unterschiedliche Modalisierungsverhalten einzelner Verben.

	<i>begründen</i>				<i>erteilen</i>			
	GK ²⁹	TK 1	TK 2	TK 3	GK	TK 1	TK 2	TK 3
insgesamt	351	84	163	104	238	145	82	11
einfaches Verb	224	27	131	66	102	49	51	2
explizit modalisiert	127	57	32	38	136	96	31	9
in Verbindung mit								
<i>dürfen</i>	2	1		1	21	16	4	1
<i>können</i>	27	4	13	10	32	24	7	
<i>lassen</i>	4		2	2	1	1		
<i>müssen</i>	11	3	4	4	4		3	1
<i>sollen</i>	12	8	1	3	4	4		
<i>wollen</i>	2		1	1				
<i>mögen</i>	1			1				
<i>vermögen</i>	6		4	2				
zwei Modalverben					1	1		
<i>sein/ haben</i> + Infinitiv	44	36	2	6	41	33	9	2
Modalpartizip	1			1	1	1		
weitere Modallexeme	17	5	5	7	29	17	8	5

Bei der Untersuchung der eingegangenen Verknüpfungen ergeben sich weitere feinkörnige Regelmäßigkeiten: So weisen in den untersuchten Korpus-texten sämtliche Belege, in denen *begründen* mit dem modalen Infinitiv verwendet wird, die Bedeutung „Gründe, etwas als Grund für etwas angeben“ (Duden Online-Wörterbuch, letzter Zugriff im Oktober 2016) auf, wie sie etwa in Kombination mit den Substantiven *Urteil*, *Beschluss* und *Entscheidung* aktualisiert wird. Neben der angeführten Lesart wird das Verb hier vor allem in der Bedeutungsvariante „den Grund

²⁸ Zu denken ist hier an die Formulierung von direktiven Sprechhandlungen im Indikativ Präsens, wie sie in Gesetzestexten anzutreffen sind, sich aber auch in der Umgangssprache finden (z.B. in der Äußerung: *Es wird gegessen, was auf den Tisch kommt*).

²⁹ GK = Gesamtkorpus, TK 1 = Gesetzestexte, TK 2 = Gerichtsentscheidungen, TK 3 = Gesetzeskommentare.

zu etwas legen, eine Grundlage für etwas schaffen“ (ebenda) verwendet, z.B. in Verbindung mit den Termini *Ermessen* und *Anspruch*.³⁰

Die Aktualisierung unterschiedlicher Bedeutungsvarianten des Verbs verweist noch einmal auf die stark vereindeutigend bzw. kontextualisierend wirkende Beziehung zwischen Verb und Nomen, die bereits weiter oben anhand der verschiedenen Bedeutungen des Verbs *spielen* zur Sprache gekommen ist. Dabei ist festzuhalten, dass die semantische Rolle von Substantiven und ihre Zugehörigkeit zu einer bestimmten semantischen Klasse einen durchaus unterschiedlichen Grad an vereindeutigender bzw. kontextualisierender Wirkung für Verben besitzen. Im Fall des Verbs *begründen* gelingt es nicht immer, mit ihrer Hilfe die beiden oben genannten Lesarten klar voneinander abzugrenzen, vielmehr kommt es zu Überschneidungen.³¹ Weitaus stärker kommt die semantische Valenz hingegen bei den verschiedenen Verknüpfungen zum Tragen, die *erteilen* als Funktions- bzw. Stützverb mit Substantiven eingeht: Als Beispiel für Nomina, die als Füllung der Akkusativergänzung dienen, sei die semantische Klasse „Aufenthaltstitel“ genannt, die neben dem Oberbegriff auch die Substantive *Aufenthaltserlaubnis*, *Aufenthaltsgenehmigung*, *Duldung*, *Visum* umfasst. Weitere Reihenbildungen sind u.a. *Befehl*, *Anordnung*, *Auftrag*, *Weisung* oder *Erlaubnis*, *Genehmigung*, *Gestattung*, *Zulassung*, *Zustimmung* (samt den dazugehörigen Komposita). Dass *erteilen* an der Bildung semantisch sehr unterschiedlicher Nominalisierungsverbgefüge beteiligt ist, wird besonders deutlich, wenn man den engeren Bereich fachsprachlicher Kommunikation verlässt und weitere allgemeinsprachliche Beispiele wie *eine Abfuhr*, *seinen Segen* oder *einen Rat erteilen* hinzuzieht.

30 Die unterschiedliche Modellierung und Ausdifferenzierung der zu unterscheidenden begrifflich-semantischen Bedeutungsverhältnisse tritt auch bei diesem Verb typischerweise erst im zwei- bzw. mehrsprachigen Vergleich voll zutage. Während in den gängigen einsprachigen Wörterbüchern zwei bis drei Bedeutungen des Verbs unterschieden werden, gehören zu den Übersetzungsmöglichkeiten im Italienischen neben *motivare*, *fondare* und *costituire* etwa *giustificare*, *istituire* sowie eine ganze Bandbreite von Substantiv-Verb-Verbindungen (*formire + spiegazione*, *esporre + motivi*, *illustrare + motivazioni*, *illustrare + ragione*, *avere + giustificazione* oder, als weitere Variante, *spiegare + il perché*). Ähnliches gilt für *erteilen*, für das im Allgemeinen eine semantisch nicht weiter ausdifferenzierte Bedeutung angegeben wird, sich aber mit Hilfe von Paralleltexten ebenfalls leicht eine längere Liste von Entsprechungen zusammenstellen lässt: neben *dare* und *impartire* kommen als weitere Wiedergabemöglichkeiten u.a. die Verben *accordare*, *assegnare*, *concedere*, *dotare* und *trasmettere* in Betracht. Auch hier erheben die angegebenen Entsprechungen keinen Anspruch auf Vollständigkeit.

31 Dies äußert sich auch darin, dass die Verknüpfung des Verbs *begründen* mit *Entscheidung* als PATIENS allein nicht ausreichend ist, um die Wiedergabe im Italienischen u.a. mit Hilfe der Kombinationen *decisione + motivare* oder *decisione + fondare* zu erklären. Die Klärung dieser Frage gehört wie ganz allgemein die Untersuchung der Kontextabhängigkeit der verschiedenen Übersetzungsvarianten zu den Aufgaben zukünftiger Forschung.

Da im Fall von Funktions- bzw. Stützverben das prädikative Nomen besonders stark vereindeutigend bzw. kontextualisierend wirkt, ist es empfehlenswert, eine Veränderung der Blickrichtung in Betracht zu ziehen und bei der Korpusrecherche nicht nur vom Verb, sondern darüber hinaus vom prädikativen Nomen auszugehen. Dadurch gelingt es, über inhaltlich eng zusammenhängende Belege Teilausschnitte des entsprechenden semantisch-lexikalischen Netzes näher zu erfassen. Zu den Belegreihen der verbalen Ausdrücke, die in den untersuchten verwaltungsrechtlichen Texten zusammen mit dem Nomen *Genehmigung* verwendet werden, gehört beispielsweise: *eine ~ befristen, erteilen, treffen, versagen, widerrufen, zurücknehmen*. Die hier auftretenden Verknüpfungen von Verb und Substantiv können ihrerseits wieder den Ausgangspunkt für eine weitere sehr feinmaschige Erfassung von Sprachgebrauchsregelmäßigkeiten werden.³² Untersucht man beispielsweise das Modalisierungsverhalten der beiden Konstrukte *Genehmigung + erteilen* und *Auskunft + erteilen*, die einen Großteil der mit *erteilen* gebildeten Nominalisierungsverbgefüge stellen, so zeichnen sich einige deutliche Gebrauchstendenzen ab: Während sich in den untersuchten Korpus-texten für *Genehmigung + erteilen* allein die unpersönliche Gebrauchsvariante des modalen Infinitivs (*'sein + zu + Infinitiv'*) findet, herrscht bei *Auskunft + erteilen* der modale Infinitiv vor, bei dem der AGENS benannt wird (*'haben + zu + Infinitiv'*). Des Weiteren verhalten sich beide Nominalisierungsverbgefüge hinsichtlich der Modalisierung durch lexikalische Ersatzformen (wie beispielsweise *berechtigt, verpflichtet, vorgeschrieben sein*) sehr unterschiedlich. Tritt bei *Genehmigung + erteilen* die Modalisierung durch weitere Modallexeme so gut wie nicht auf, ist sie bei *Auskunft + erteilen* relativ häufig anzutreffen (s. dazu Beispiel 9):

- (9) Bund und Länder sind **verpflichtet**, sich gegenseitig alle **Auskünfte zu erteilen**, die zur Durchführung der Aufgaben der Raumordnung notwendig sind. (§ 19 Abs. 3 ROG)

Angesichts der niedrigen Belegzahlen, die sich bei der Verwendung mehrerer Auswahlkriterien ergeben – worauf gleich noch näher einzugehen ist –, wurden beide Gebrauchstendenzen anhand eines ca. 1 Mio. Tokens großen Korpus aus verwaltungsrechtlichen Gesetzestexten überprüft. Dabei bestätigten sich beide Gebrauchstendenzen nicht nur, sondern es kristallisierte sich zudem heraus,

³² Die automatische Erstellung von Gebrauchsprofilen bzw. von Word Sketches bei erweiterten lexikalischen Einheiten (Nominalisierungsverbgefügen wie den angeführten deutschen und italienischen Konstrukten, aber auch Komposita, Zwillingformen u.a.m.) oder weiteren Arten der Wortkombinationen ist noch nicht ausgereift und wird beispielsweise vom DWDS-Wortprofil bisher noch nicht oder nur in Ansätzen angeboten (s. dazu v. a. Kilgariff et al. 2012).

dass für das Konstrukt *Auskunft + erteilen* – in dem durch die Textauswahl umrissenen kommunikativen Teilbereich („Handlungsraum“) – u.a. die Verwendung der Ersatzform *verpflichtet sein* als charakteristisch gelten kann.

5.2 Wortkombinationen

Die starke Reduzierung der Belegzahlen, die sich allein durch die Verknüpfung mehrerer Suchwörter ergibt, verweist auf die kontextualisierende Wirkkraft, die sinnhaft verknüpfte Wörter bzw. Wortbildungselemente als solche entfalten. Wie anhand größerer Textkorpora leicht veranschaulicht werden kann und wie bereits oben deutlich wurde, schrumpft die Anzahl der Treffer regelrecht zusammen, wenn nach bestimmten Kombinationen gesucht wird (s. dazu vor allem Tabelle 4, aber auch Tabelle 3).

Tabelle 4: Suchanfragen im Vergleich: Reduzierung der Belegzahlen durch die Suche nach Wortkombinationen (Concgrams).

Korpusbestände ³³	<i>erteilen</i>	<i>erteilen + Auskunft</i>	<i>erteilen+ verpflichten</i>	<i>erteilen + verpflichten + Auskunft</i> ³⁴
DWDS-Korpora	64073	2517	247	68
deTenTen13	467959	42037	4472	1729 (1513)
TK1 +	452	146	22	17 (16)

Beobachtungen dieser Art führen notwendig zur Forderung nach möglichst umfangreichen Korpusbeständen.³⁵ Umgekehrt kann die selektierende Wirkung, die von der Kombination lexikalischer Einheiten ausgeht, dazu genutzt werden,

³³ Bei den DWDS-Korpora wurde die Default-Einstellung der Korpuswerkzeuge verwendet, bei der Wortkombinationen jeweils innerhalb eines Satzes erfasst werden. Die Angaben beziehen die aus urheberrechtlichen Gründen nicht anzeigbaren Korpusbelege mit ein. Bei den anderen beiden mit Hilfe des Tools Sketch Engine untersuchten Korpora (deTenTen13 und TK1, dem auf ca. 1 Mio Tokens aufgestockten Subkorpus aus Gesetzestexten des deutschen Verwaltungsrechts) wurde ebenfalls eine Lemma-Suche durchgeführt, zu diesem Zweck jedoch eine Kollokationsspanne von ± 15 gewählt. Die bei der Kombination *erteilen + verpflichten + Auskunft* in runden Klammern hinzugefügte Belegzahl bezieht sich auf das Vorkommen im selben Satz. Die Wildcardsuche konnte nur für die zuletzt genannten Korpora durchgeführt werden.

³⁴ In Prozentwerten ausgedrückt liegt die Anzahl für das Concgramm *erteilen + verpflichten + Auskunft* in den beiden Großkorpora bei unter 0,4 % und in dem Korpus aus Gesetzestexten bei unter 4% aller erfassten Belege des Verbs *erteilen*.

³⁵ Siehe dazu u.a. Atkins/Rundell, die sich jedoch nicht nur auf Wortkombinationen als solche, sondern allgemein auf die Bedeutungsdifferenzierung von Lexemen beziehen. Sie führen aus: „A lemma like the verb *break*, with almost 19,000 hits in the BNC [British National Corpus, D.H.]

die Aktualisierung einzelner Bedeutungen, mikroskopisch feine Regelmäßigkeiten im Sprachgebrauch und die Ausgestaltung kommunikativer Aufgaben gezielt zu untersuchen. Zur Veranschaulichung sei das letzte Beispiel noch einmal aufgegriffen. Anhand des Vorkommens der Belege für die Wortkombination *erteilen + verpflichten + Auskunft* kann aufgezeigt werden, dass die Verwendung sogenannter Concgramme³⁶ hier neue Untersuchungsmöglichkeiten eröffnet.

Die im German Web 2013 (deTenTen13) vorkommende Anzahl von über 1500 Belegen für diese Wortkombination erlaubt eine genauere Untersuchung der sich hier abzeichnenden Sprachgebrauchstendenzen. Dabei ist bereits die Beobachtung interessant, dass der Anteil der Belege, in denen die drei Wörter nicht direkt aufeinander bezogen sind, bei unter 5% liegt. Zudem bestätigt sich von Neuem, dass die Verknüpfung des Konstrukts *Auskunft + erteilen* mit *verpflichtet sein* als besonders typisch gelten kann. Allein auf die Realisierungsform *ist/sind verpflichtet ...* entfallen über zwei Drittel aller Belege. Dabei herrscht – wie im verwaltungsrechtlichen Korpus aus Gesetzestexten, bei dem mit zwei Ausnahmen alle Belege diese syntaktische Struktur aufweisen – der Gebrauch des Indikativ Präsens mit einem von ihm abhängigen Infinitiv (wie in Beispiel 9) vor. Neben der Gebrauchsvariante *verpflichtet sein*, die mit ca. 80% mit Abstand den Hauptanteil aller Belege im deTenTen13 ausmacht, lassen sich hier drei weitere Gebrauchsweisen deutlich voneinander abgrenzen: *sich verpflichten*, *jemanden verpflichten* und *verpflichtet werden*, deren jeweiliger Anteil zwischen 4 und 6% liegt. Andere Verwendungsweisen (wie *sich verpflichtet fühlen* oder

appears to be well supplied with data. But there is a strong correlation between a word's frequency and its complexity. Thus *break* has at least twenty different meanings and up to a dozen phrasal verbs (some of them polysemous). It also participates in innumerable phrases, grammatical patterns, and lexical collocations. And just as some words are frequent and some rare, the same applies to meanings, phrases, patterns, and collocations. Consequently, for some uses of *break*, we may find the evidence surprisingly thin. When we find that the BNC has only eight examples of the combination 'break someone's serve/service' (in tennis), our 19,000 hits no longer look so impressive. [...] We don't actually know how much data we need in order to account for a given linguistic feature, be it a word, a meaning, or a word combination. What we do know is that the more data we have, the more we learn." (Atkins/Rundell 2008: 60-61).

36 Concgramme sind "sets of words that co-occur regardless of constituency variation (e.g. AB and A*B), positional variation (e.g. AB and BA), or both" (Cheng et al. 2009, 236), die bei einer bewussten Auswahl der Suchwörter als Kombination zweier oder mehrerer KWIC-Recherchen begriffen werden können. Die Möglichkeit, Kookkurrenzen unabhängig von ihrer linearen Abfolge zu erfassen (s. dazu außer dem bereits genannten Beitrag vor allem Cheng et al. 2006 und Greaves 2009), ist gerade bei der Untersuchung deutschsprachiger Texte von Vorteil, da das Deutsche als Klammersprache wesentlich durch das Vorkommen diskontinuierlicher Konstituenten geprägt ist. Der Einsatz von Concgrammen stellt aber auch sonst für die korpuslinguistische Untersuchung komplexer Texte ganz allgemein einen entscheidenden Fortschritt dar.

sehen) sind nur in wenigen Einzelfällen belegt. Der Umstand, dass die weiteren im deTenTen13 belegten Gebrauchsvarianten mit einer einzigen Ausnahme nicht im verwaltungsrechtlichen Korpus aus Gesetzestexten vorzufinden sind, verweist auf die jeweils erfüllten kommunikativen Aufgaben der Sprachmittel und ist im Wesentlichen damit zu erklären, dass sie anderen Handlungszusammenhängen zuzuordnen sind. So dürfte beispielsweise die Verwendungsweise *sich verpflichten* eher in privatrechtlichen Verträgen oder internationalen Abkommen eine Rolle spielen (Beispiele 10 und 11).³⁷

- (10) Der Käufer **verpflichtet sich**, uns hierfür sämtliche **Auskünfte** zu **erteilen** und die erforderlichen Unterlagen dafür auszuhändigen.
- (11) In dem neuen DBA **verpflichtet sich** die Schweiz, auch dann **Auskünfte** zu **erteilen**, wenn kein Fall von Steuerbetrug nach schweizerischem Recht vorliegt.

Zu den weiteren Gebrauchstendenzen, die sich abzeichnen, gehört, dass die explizite Modalisierung der Gebrauchsvariante *verpflichtet sein* bei unter 2% liegt und im Wesentlichen auf den Gebrauch der Modalverben *können* und *sollen* beschränkt ist. Bei den Gebrauchsweisen *jemanden verpflichten* und *verpflichtet werden* liegt der Anteil der explizit modalisierten Belege hingegen bei ca. 30 % und weist darüber hinaus ein ungleich größeres Spektrum an Modalverben auf. Auch hier spiegelt der Sprachgebrauch die kommunikativen Aufgaben der Äußerungen wider, in denen es u.a. um die rechtliche Klärung der Auskunftspflicht geht oder beispielsweise politische Standpunkte vertreten werden (Beispiele 12 und 13).

- (12) In einem aktuellen Urteil hat das Verwaltungsgericht Frankfurt/Main (AZ: 1 K 3874/08.F(2)) entschieden, dass die Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin) einen Rechtsanwalt dazu **verpflichten darf, Auskünfte** zu **erteilen**.
- (13) Besonders schmackhaft ist die Einführung des Blockwartes durch das Gesetz: Darüber hinaus **will** es Wohnungseigentümer beziehungsweise Vermieter dazu **verpflichten, Auskunft** über die bei ihm wohnenden Personen zu **erteilen**.

Die mit Hilfe der Wortkombination *verpflichten + erteilen + Auskunft* erfassten Korpusbelege lassen zudem eine zügige computergestützte Erfassung inhaltlich

³⁷ Die folgenden Beispielbelege (12-19) stammen aus dem deTenTen-Korpus13 (<https://the.sketchengine.co.uk>).

relevanter Unterscheidungen und Bestimmungen bzw. ihrer Versprachlichung zu, so dass die Ausgestaltung des betreffenden Ausschnitts aus dem semantisch-lexikalischen Netz besonders feinkörnig nachgezeichnet werden kann. Zu nennen sind hier beispielsweise die rechtlichen Grundlagen der Auskunftspflicht, die sich an den erfassten Belegen ablesen lassen, etwa wenn das Concgramm um die Wörter *gesetzlich* oder *vertraglich* oder um die Nennung bestimmter Normen erweitert wird. Dabei stellen diese Spezifizierungen zugleich einen wichtigen Schlüssel für das Verständnis der zahlenmäßig überwiegenden Belege dar, in denen diese Information mitgemeint ist (Beispiele 14-19).

- (14) Die Gemeinden und die Staatsdienste sind **verpflichtet**, unentgeltlich die zur Abklärung nötigen **Auskünfte** zu **erteilen**.
- (15) Was viele nicht wissen: Das Finanzamt ist **gesetzlich verpflichtet**, bei Fragen zur Lohnsteuer **Auskünfte** zu **erteilen**.
- (16) **Auskunft erteilt** die SCHUFA nur, wenn sie **vom Gesetz her** dazu **verpflichtet** ist.
- (17) **Nach § 2314 BGB** ist der Erbe **verpflichtet**, einem Pflichtteilsberechtigten **Auskunft** über den gesamten Nachlass zu **erteilen**.
- (18) **Gemäß § 97 InsO** sind Sie **verpflichtet**, **Auskunft** über Ihr gesamtes Vermögen dem Treuhänder zu **erteilen**.
- (19) Darüber hinaus sei erneut zu betonen, dass die Beklagte nunmehr auf **aufgrund der Richtlinie 2004/48/EG**, dort Art. 8 Abs. 1, **verpflichtet** wäre, **Auskunft** zu **erteilen**.

...

Überhaupt stellt die Wortkombination eine robuste Grundlage für die weitere Erfassung von Sprachmustern sowie den damit verbundenen inhaltlichen Ausdifferenzierungen dar. Über das Concgramm *verpflichten + erteilen + Auskunft* kann beispielsweise erfasst werden, dass das Reziprok-Pronomen *einander* in diesem Zusammenhang im Wesentlichen zusammen mit Verwandten, Ehegatten und Hinterbliebenen bzw. Miterben Verwendung findet und die Erteilung von Auskünften vielfach erst einzufordern ist, worauf der häufig verwendete Zusatz *auf Verlangen* hinweist (Beispiel 20).

- (20) § 1580 BGB bestimmt, dass geschiedene **Ehegatten einander verpflichtet** sind, **auf Verlangen** über ihre Einkünfte und ihr Vermögen **Auskunft** zu **erteilen**.

Durch die Zusammenschau verschiedener Belege tritt schließlich u.a. auch der teilweise nahezu zwillingsformhafte Charakter der Verknüpfung von *berechtigt und verpflichtet* zutage. Dabei ist festzuhalten, dass die beiden Verben in allen 26 Belegen in einem direkten Bezug zueinanderstehen und in der Mehrzahl der Fälle dasselbe Subjekt haben (Beispiel 21).

- (21) Das Concgramm *erteilen + verpflichten + Auskunft + berechtigt*
- a. Der Bürgermeister hat beim Rechnungsprüfungsverfahren kein Stimmrecht; er ist jedoch **berechtigt** und **verpflichtet**, **Auskünfte** zu **erteilen**.
 - b. Inwieweit ist das Einwohnermeldeamt **berechtigt** bzw. **verpflichtet**, **Auskunft** über die **Adresse** des Frauenhauses zu **erteilen**?
 - c. AGENTOURS ist weder **berechtigt**, noch **verpflichtet**, diesbezügliche **Auskünfte** zu **erteilen** oder Bescheinigungen zu erstellen.
 - d. Der Betriebsleiter nimmt an den Sitzungen des Betriebsausschusses mit beratender Stimme teil; er ist **berechtigt** und auf Verlangen **verpflichtet**, zu den Beratungsgegenständen Stellung zu nehmen und **Auskünfte** zu **erteilen**.
 - e. Im vorliegenden Fall wäre die Antragsgegnerin **berechtigt** und bei Vorliegen der Voraussetzungen des § 101 Abs. 2 UrhG auch **verpflichtet**, die von der Antragstellerin begehrte **Auskunft** zu **erteilen**.
 - f. Ein webbasiertes Auskunftstool hilft Ihnen, Auskunftsanfragen bezüglich Studierendendaten von Behörden, Krankenkassen, Ermittlungsbehörden u.v.a. Stellen kompetent und schnell dahingehend zu beurteilen, ob Sie **berechtigt** oder gar **verpflichtet** sind, **Auskünfte** zu **erteilen**.
 - g. i-potentials ist hierbei **berechtigt**, in bestimmten Einzelfällen **Auskunft** zu **erteilen** oder ist aufgrund gesetzlicher Vorschriften zu einer derartigen **Auskunftserteilung** gegenüber den zuständigen Stellen **verpflichtet**.

...

Der zuletzt angeführte Beispielbeleg verweist dabei auf das Potential, das die Verwendung von Wildcards beim Einsatz von Concgrammen besitzt. Da mit ihrer Hilfe auch Komposita und Derivata automatisch erfasst werden können, kann den für das Deutsche typischen Wortbildungsmechanismen besser Rechnung getragen werden, was sich unmittelbar in einem deutlichen Anstieg der Trefferquoten niederschlägt (s. Tabelle 5).

Die Offenheit bzw. Unterdeterminiertheit der syntagmatischen Bezüge, die den Einsatz von Concgrammen kennzeichnet, wird durch den Einsatz von

Tabelle 5: Suchanfragen im Vergleich: Trefferquoten bei der Suche nach Conccgrammen mit und ohne Wildcards.

Korpusbestände ³⁸	<i>erteilen</i>	<i>erteilen + Auskunft</i>	<i>erteilen + verpflichten</i>	<i>erteilen + verpflichten + Auskunft</i>
TK1 +	452	146	22	17
deTenTen13	467959	42037	4472	1729
mit Wildcardsuche (*erteil*, *pflicht*, *Ausk*)				
TK1 +	913	221	83	50
deTenTen13	3324804	69296	41755	6852 ³⁹

Wildcards noch weiter verstärkt. Wie die angeführten Belege zeigen, können mit Hilfe dieser Suchstrategie die verschiedenen vorkommenden Verknüpfungen bzw. bestehenden Ausdrucksmöglichkeiten erfasst, aufgezeigt und – für NutzerInnen lexikographischer Nachschlagewerke besonders relevant – leichter nachgebildet werden. Durch die Feinmaschigkeit ihrer Erfassung wird zudem die Voraussagbarkeit ihres Auftretens erhöht.

(22) Das Conccgramm **erteil* + *pflicht* + *ausk**

(unter besonderer Berücksichtigung von Komposita e Derivata)

- a. Die Fallkonstellation, die Herr Vetter in seinem Beitrag zum Gegenstand macht, nämlich die „einfachen Anfragen der Polizei, welche Person hinter einem Account steht“, betrifft die **Verpflichtung** des TK-Anbieters nach § 113 TKG, **Auskunft** über die Daten zu **erteilen**, die nach §§ 95, 111 TKG erhoben wurden.
- b. Jedoch darf eine **Auskunft**, die der Schuldner gemäß seiner **Verpflichtung** nach Satz 1 **erteilt**, in einem Strafverfahren oder in einem Verfahren nach dem Gesetz über Ordnungswidrigkeiten gegen den Schuldner oder einen in § 52 Abs. 1 der Strafprozeßordnung

³⁸ Die Suchergebnisse beziehen sich wie gehabt auf eine Kollokationsspanne von ± 15 .

³⁹ Es handelt sich um Rohdaten, bei denen fehlerhafte Treffer nicht aussortiert wurden. Doch wurde überprüft, inwieweit sich die durch die Wildcardsuche induzierten Fehler auf die Suchergebnisse auswirken. Ihre Anzahl ist bei dem untersuchten Beispiel erstaunlich niedrig: Eine Gegenprobe ergab, dass bei der Dreierkombination weniger als 20 Belege betroffen sind. Schwerer ins Gewicht dürften Treffer fallen, die wie im folgenden Beleg in keinem inhaltlichen Zusammenhang mit dem Problem der gesetzlichen oder vertraglichen Auskunftspflicht stehen: „Gegenstand der Leistungen von WELCOME MEDIA GMBH ist die Bereitstellung einer telefonischen Vermittlungsplattform zur **Auskunftserteilung** bzw. die Vermittlung von Kontakten zwischen Kunden und Beratern über gebührenpflichtige Rufnummern per Telefon.“ (deTenTen13).

- bezeichneten Angehörigen des Schuldners nur mit Zustimmung des Schuldners verwendet werden.
- c. Alle Organe und Gliederungen des Bezirksverbandes sind der Kontrollkommission zur **Auskunftserteilung verpflichtet**.
 - d. Diese Beschlüsse hob der Senat auf und begründete dies damit, dass die **Verpflichtung** zur **Auskunftserteilung** auf die Übermittlung vorhandener Datenbestände beschränkt sei.
 - e. Was jedoch immer enthalten ist, ist der Eindruck, dass die **Auskunftserteilung Pflicht** wäre – was allerdings nicht immer stimmen muss...
 - f. Ein Auskunftsvertrag mit einer vertraglichen **Hauptpflicht** zur **Auskunftserteilung** kommt nur bei einer entsprechenden ausdrücklichen Vereinbarung zustande.
 - g. Der zur **Erteilung** einer **Auskunft Verpflichtete** kann die Auskunft auf solche Fragen verweigern, deren Beantwortung ihn selbst oder einen der in § 383 Absatz 1 Nr. 1 bis 3 der Zivilprozessordnung bezeichneten Angehörigen der Gefahr der Verfolgung wegen einer Straftat oder eines Verfahrens nach dem Gesetz über Ordnungswidrigkeiten aussetzen würde.
 - h. Die Finanzbehörde kann verlangen, daß der **Auskunftspflichtige** schriftlich **Auskunft erteilt**, wenn dies sachdienlich ist.
 - i. Früher musste der auskunftsverpflichtete Ehegatte nur Auskunft über den Bestand seines Endvermögens zum Stichtag erteilen.
 - j. Die vom Erben erbetene **Auskunft** sollte vom **Auskunftsverpflichteten** unverzüglich und mit gebotener Sorgfalt in schriftlicher Form **erteilt** werden.
 - k. 1 Das Landesamt für Verfassungsschutz **erteilt** dem Betroffenen auf Antrag kostenfrei **Auskunft** über die zu seiner Person in Dateien oder Akten gespeicherten Daten. 2 Die **Auskunftsverpflichtung** besteht nur, soweit der Betroffene ein besonderes Interesse an einer Auskunft darlegt.
- ...

Darüber hinaus zeigt sich erneut, dass Concgramme durch die mögliche Zusammenschau von Ausdrucks- bzw. Formulierungsvarianten einen methodischen Zugang zur systematischeren Erfassung und Untersuchung von lexikalischen Varianten einschließlich nicht realisierter Äußerungskomponenten zu eröffnen scheinen. Dass bei *Auskunftsverpflichtung* und *Auskunftspflicht* das Verb *erteilen* oder das von ihm abgeleitete Verbalsubstantiv *Erteilung* ausgespart ist, kann beispielsweise leicht aus den verschiedenen verwandten Verwendungskontexten erschlossen werden bzw. wird von diesen nahegelegt.

Die stark kontextualisierende Wirkung von Wortkombinationen lässt sich bereits an der Häufigkeitsverteilung der Lexeme ablesen, die gemeinsam in abgeschlossenen Sinneinheiten wie z.B. vollständigen Sätzen vorkommen. Wort- bzw. Lemmatalisten von Sätzen, die ein bestimmtes Concgramm enthalten, stellen dementsprechend eine relativ leicht umzusetzende Möglichkeit dar, die Kontextualisierungswirkung von Wortkombinationen zur gezielten Erfassung bestimmter Sachzusammenhänge⁴⁰ und den zu ihrer Versprachlichung verwendeten Ausdrucksmitteln einzusetzen. So zeigt sich, dass in den Lemmatalisten, die auf der Grundlage des Concgramms *erteilen + verpflichten + Auskunft* erstellt wurden, schwerpunktmäßig Lexeme enthalten sind, die auf das Erfordernis bestimmter Auskünfte abheben, und zwar sowohl im Fall des ca. 1 Mio. Tokens umfassenden Teilkorpus aus verwaltungsrechtlichen Gesetzestexten als auch beim ca. 19,9 Mrd. Tokens großen Webkorpus deTenTen13. Dieser thematische Schwerpunkt, der sich in den Links- und Rechtserweiterungen des Lemmas *Auskunft* (Adjektivattribute und Relativsätze) widerspiegelt, tritt besonders deutlich hervor, wenn der Blick auf andere Elemente der semantische Klasse und der Wortfamilie der betreffenden Lexeme ausgeweitet wird (in unserem Fallbeispiel sind neben den besonders häufig vorkommenden Adjektiven *erforderlich* und *notwendig* insbesondere *nötig*, *erfordern*, *benötigen* zu nennen). Bei den entsprechenden Informationen des Wortprofils bzw. der Word Sketches zu *Auskunft*, die auf der nicht weiter gefilterten Gesamtmenge aller Konkordanzzeilen beruhen, kristallisiert sich dieser Themenschwerpunkt hingegen nicht heraus. Zur

Tabelle 6: Word Sketches im Vergleich.

Korpora / Datenauswahl	Modifikatoren des Lexems <i>Auskunft</i> (Inhaltswörter)
I. Sämtliche Belege des Suchwortes <i>Auskunft</i>	
Teilkorpus Gesetzestexte deTenTen13	<i>erforderlich, notwendig, schriftlich, unverzüglich begehrt, detailliert, eingeholt, erbeten, erforderlich, erschöpfend, erteilt, fachkundig, kompetent, mündlich, rechtsverbindlich, schriftlich, unentgeltlich, unrichtig, verbindlich, verlangt, verlässlich, wahrheitsgemäß, weitergehend</i>
II. Belege des Concgramms <i>erteilen + verpflichten + Auskunft</i>	
deTenTen13	<i>begehrt, (ihrer öffentlichen Aufgabe) dienend, entsprechend, erforderlich, gewünscht, nötig, notwendig, schriftlich, umfassend, unverzüglich, verbindlich, verlangt, vollständig, wahrheitsgemäß</i>
Teilkorpus Gesetzestexte	<i>erforderlich, notwendig</i>

⁴⁰ Zum diskursanalytischen Potenzial korpuslinguistischer Untersuchungsmöglichkeiten vgl. stellvertretend Felder 2012.

Veranschaulichung werden in Tabelle 6 die jeweils angegebenen Modifikatoren (zumeist attributiv oder prädikativ verwendete Adjektive) aufgeführt, wobei zu beachten ist, dass bei der automatischen Generierung von Wortprofilen bzw. Word Sketches aufgrund der gewählten Voreinstellungen im Allgemeinen lediglich häufigere Verbindungen angezeigt werden.

Der Vergleich der verschiedenen Angaben zeigt darüber hinaus, wie vorteilhaft bei Suchanfragen dieser Art die Verwendung möglichst umfangreicher Korpora ist und dass im Fall von Lexemen, die der Allgemeinsprache oder allgemeinen Wissenschaftssprache zuzurechnen sind, Großkorpora sehr gute Dienste leisten können. Ein genauerer Blick auf die Beispielbelege macht jedoch zugleich deutlich, dass die Qualität der Suchergebnisse durch einen Datenzuschnitt auf der Grundlage metasprachlicher Informationen zum Sprachregister, zum Rechtsgebiet, zur Zugehörigkeit zu einer bestimmten Rechtsordnung bzw. Nationalvarietät des Deutschen, zur Textart u.a. noch erheblich verbessert würden.

Wie abschließend an einem weiteren Beispiel gezeigt werden soll, kann die stark kontextualisierende Wirkung von Wortkombinationen auch für das Auffinden von Ausdrucksmöglichkeiten bzw. Übersetzungen genutzt werden. Als primäre Datengrundlage dient ein dreisprachiges Korpus (DE, EN, IT) aus Texten zur Tourismuskommunikation.⁴¹ Auch hier zeigt sich, dass bereits die Lemmataliste der Sätze des deutschen Teilkorpus, in denen die Wortkombination *Mensch + Landschaft*⁴² vorkommt, über die Themen Auskunft gibt, die in diesem so eingegrenzten Kommunikationszusammenhang bevorzugt zur Sprache kommen. Dazu gehört mit an erster Stelle der Einfluss des Menschen auf die Landschaft, zu dessen Versprachlichung u.a. das Verb *prägen* gebraucht wird (Beispielreihe 23).

(23) Das Concogramm *Mensch + Landschaft + prägen*

- a. Wie die **Menschen** die Kultur und **Landschaft** rund um den Nationalpark **prägten**, erfahren Sie bei einem Spaziergang durch Berchtesgaden.

⁴¹ Der Grundstock zu dem dreisprachigen, mittlerweile für das Deutsche und Englische jeweils ca. 2 Mio. Tokens umfassenden Korpus (DE – IT – EN) wurde im Rahmen des bereits erwähnten, von der Regione Sardegna geförderten Forschungsprojekts zur quantitativ-qualitativen Untersuchung authentischer Textvorkommen (2010-2012) gelegt. Die Texte wurden vorwiegend den Webauftritten deutscher, englischer und italienischer Nationalparks entnommen. Mein besonderer Dank gilt A. Lombardi für die Möglichkeit, die italienische Datengrundlage mit Hilfe des im Rahmen des TouriTerm-Projekts (Lombardi 2013) zusammengestellten Teilkorpus aus Tourismustexten zu Brescia und Umgebung auf ca. 1 Mio. Tokens zu erhöhen.

⁴² Beide Wörter gehören zu den häufigsten Lexemen in dem deutschen Teilkorpus und wurden schon aufgrund der dadurch relativ guten Datengrundlage als Beispiel ausgewählt.

- b. Die **Landschaft** des Naturparks wurde durch die Tätigkeit des **Menschen** entscheidend **geprägt** und deshalb sind die über Jahrhunderte entstandenen Dörfer und Siedlungen ein unverzichtbarer Teil der Kulturlandschaft.
- c. Der **Mensch** betreibt Land- und Forstwirtschaft, Jagd und Bergbau, er rodet und erholt sich und **prägt** damit das Gesicht der **Landschaft**

Über die Suche der um eines der Wörter gekürzten Kombinationen, z.B. *Mensch- + Landschaft*, können weitere lexikalische Varianten erfasst werden. Dabei erweist sich auch hier wieder, wie hilfreich es für die Erfassung der gesuchten Ausdrucksmittel ist, Lexeme anhand von semantischen Klassen und Wortfamilien zusammenzufassen und mit Hilfe von Gebrauchsprofilen zu ordnen.

(24) Das Concogramm *Mensch + Landschaft*

- a. Zur Mannigfaltigkeit natürlicher Umweltfaktoren kommen infolge der massiven **Einwirkungen** des **Menschen** auf diese **Landschaft** noch eine – bisher wenig bearbeitete – Fülle anthropogener Faktoren, die alle zusammen letztlich Besiedlung, Bestandeshöhe, Verbreitungsmuster und Entwicklungstrend der Tierpopulationen bestimmen.
- b. Welche Spuren hat der **Mensch** durch seinen Jahrhunderte lange [sic!] **Einfluss** auf Natur und **Landschaft** hinterlassen?
- c. Es sollten Grundlageninformationen zur Beurteilung der Art in unserer stark vom **Menschen beeinflussten Landschaft** beschafft werden.
- d. Die **Alpenlandschaft** ist sehr stark vom **Wirken** des **Menschen beeinflusst**, denn er besiedelte bereits vor über 3000 Jahren ihre Hochflächen.
- e. Die **Kulturlandschaft** des Naturparks Lüneburger Heide ist über Jahrhunderte durch den **Einfluss** des **Menschen** entstanden und bietet heute einen Lebensraum für viele seltene Pflanzen- und Tierarten.

Auf die gleiche Art und Weise können Übersetzungsmöglichkeiten erfasst bzw. überprüft werden. So zeigt sich beispielsweise, dass in dem englischen Teilkorpus in diesem Zusammenhang häufiger *people* statt die wortwörtliche Wiedergabe durch *man* gebraucht wird und das Verb *shape* aufgrund seiner Vorkommenshäufigkeit in Verbindung mit der Wortkombination *people + landscape* hier zu den wichtigsten Übersetzungen des Verbs *prägen* gehört.

(25) Beispielbelege für das Concogramm *landscape + shape*

- a. It is a **landscape shaped** by **man**, by history and by the animals that still graze it today.

- b. You may think that the **landscape** of the Brecon Beacons is unspoilt and natural, but in fact **people** have **shaped** and changed it over many thousands of years.
- c. A unique **landscape** of moorland, woodland, valleys and farmland, **shaped** by **people** and nature over thousands of years.
- d. **People** have lived here for 8000 years and the **landscape** has been **shaped** by their activities.

Ähnlich können auf diese Art und Weise entsprechende italienische Ausdrucksmöglichkeiten ermittelt werden:

- (26) Das Concgramm *uomo + paesaggio*
- a. I parchi dello Strone e Mella Morta Campagne e **paesaggi modellati dall'uomo**.
 - b. **L'uomo**, costruendo per scopi idroelettrici la diga del lago artificiale di Malga Bissina terminata nel 1958, ha **ridisegnato il paesaggio** e l'ambiente originale di una parte della valle.

Bei der Suche nach diesen Wortkombinationen erweist sich die Hinzuziehung von Großkorpora als besonders hilfreich, da es die bereits zusammengestellten Teilkorpora zur Tourismuskommunikation trotz ihres vergleichsweise großen Umfangs nicht ermöglichen, eine größere Anzahl von Beispielbelegen zu erfassen und Übersetzungshypothesen anhand einer ausreichend großen Datengrundlage zu überprüfen. Wie die folgenden Belege zeigen, können die Wiedergabemöglichkeiten im Italienischen mit Hilfe des itTenTen vergleichsweise leicht ergänzt werden:

- (27) Das Concgramm *uomo + paesaggio* – weitere Beispielbelege aus dem itTenTen10
- a. Visiterete gli ambienti e i **paesaggi** che la natura e l'attività **dell'uomo** hanno **plasmato** in un territorio eccezionale: un piccolo assaggio di Parco quale invito alla scoperta di una grande Area Protetta.
 - b. Questo bel casale, perfettamente recuperato, è inserito in un **paesaggio** ideale favorito da una natura spettacolare ma anche **plasmato** dall'**uomo**.
 - c. Il **paesaggio** attuale **è il risultato di trasformazioni naturali e di modificazioni operate** dall'**uomo** tendenti ad utilizzare al massimo il terreno per le attività agricole.
 - d. Nelle zone maggiormente abitate, **l'uomo** ha **trasformato** completamente o quasi il **paesaggio** naturale vegetale.

- e. Ma l'Etna è anche un **paesaggio creato dall'uomo**, che per secoli ha cercato testardamente di opporsi alla forza del vulcano trasformando la sterilità della lava in fertili campi coltivati, piantando viti e castagni, costruendo masserie.
- f. La montagna, considerata dai cittadini un luogo ostile, lontano, oppure un semplice playground per lo sci e gli sport più o meno estremi, è in realtà un **paesaggio forgiato dall'uomo**, che in millenni di civilizzazione ha trasformato un ambiente ostile in un luogo dove abitare, lavorare, crescere i propri figli.

...

Von ganz besonderem Wert erscheint beim Gebrauch von Concgrammen schließlich der Umstand, dass – zumindest in den bisher näher untersuchten Fällen (*verpflichten + erteilen + Auskunft; Mensch + Landschaft + prägen*) – selbst die in den großen Webkorpora erfassten Belege zumeist thematisch zusammenhängen, was wesentlich auf die sie kennzeichnenden stärker ausgeprägten interlexematischen Bezüge zurückzuführen sein dürfte. Bereits durch die hierbei aktivierten, noch näher zu erforschenden Filter gelingt es, einen kommunikativen Rahmen aufrechtzuerhalten, der bei Suchanfragen, die sich lediglich auf eine lexikalische Einheit beschränken, im Allgemeinen verloren geht.

6 Schluss und Ausblick

Die wichtigsten Ergebnisse der vorliegenden Studie, die gewissermaßen ein Plädoyer für eine stärkere Beachtung und weitere Erforschung bereits bekannter sprachlicher und außersprachlicher Ordnungsprinzipien bei der Erstellung lexikographischer Ressourcen darstellt,⁴³ seien an dieser Stelle noch einmal kurz zusammengefasst:

Eine Verbindung der verschiedenen, auf außer- und innersprachlichen Organisationsprinzipien beruhenden Kontextualisierungsstrategien scheint für die Bewältigung großer Mengen von Sprachdaten besonders vorteilhaft zu sein. Durch ihre Verwendung gelingt es, eine kommunikativ motivierte Selektion von Daten vorzunehmen, die ergänzend an die Seite von Sample-Verfahren treten kann, bei denen der kommunikative Zusammenhang im Allgemeinen nicht gewahrt wird. Als besonders effektiv hat sich dabei die Verwendung von Concgrammen aus sinnhaft miteinander verknüpften Wörtern bzw. Wortelementen erwiesen, die darüber hinaus für die Erfassung von Ausdrucksmitteln und ihren Varianten besonders geeignet zu sein scheinen. Für die Erstellung von Gebrauchsprofilen bzw. Word

⁴³ In diesem Zusammenhang sei auch auf Gries (2014) verwiesen.

Sketchs, die den Wortbildungsmechanismen der deutschen Sprache besser Rechnung tragen können, wäre dabei eine Berücksichtigung der verschiedenen lexikalischen Wortbildungselemente oder von mit Wildcards versehenen Suchwörtern wünschenswert. Ebenso anstrebenwert wäre die automatische Zusammenfassung von Lexemen, die derselben semantischen Klasse zuzurechnen sind, eine bestimmte semantische Rolle erfüllen oder einer Wortfamilie angehören.

Hervorzuheben ist schließlich, dass bei nicht spezialisierten Korpora eine Datengrundlage im Milliardenbereich notwendig ist, um im Fall der angeführten Beispiele eine größere Anzahl für Belege von Dreier-Concgrammen zu erfassen und sich die Belegzahl bereits bei Vierer-Concgrammen wieder drastisch verringert. Dies bedeutet u.a., dass die Annahmen über ausreichende Datenmengen oder den Umfang repräsentativer Referenzkorpora – wie schon in der Vergangenheit mehrfach geschehen (s. dazu u.a. Kupietz/Lüngen 2014) – noch weiter nach oben hin korrigiert werden müssen. Auch im Hinblick auf interlexematische Beziehungen darf man sich durch die Erfassung und Untersuchung von Concgrammen weitere Einsichten in den Aufbau und die Funktionsweise sprachlicher Äußerungen erhoffen. Neben Fragen zur sprachlichen Variation dürfte hier insbesondere von Interesse sein, wie stark die Beziehungen zwischen den verschiedenen Wörtern bzw. Wortelelementen ausfallen und wieviel sie jeweils zur Bedeutungskonstituierung beitragen.

Anmerkung: Es handelt sich um die Ausarbeitung und Fortschreibung des im Oktober 2015 auf der Hildesheimer SELA-Konferenz gehaltenen Vortrags „Developing bi- and multilingual lexicographical tools for LSP and domain-specific communication: different strategies of contextualization“ (*Hildesheim Conference on electronic Dictionaries as Information Tools*, 7.-8. 10. 2015). Der Beitrag stellt den vorläufigen Abschluss der Untersuchungen der Verfasserin zur Polyfunktionalität als methodologischem Grundproblem der Korpusanalyse dar. Durch die Einarbeitung früherer Forschungsergebnisse, auf die in der vorliegenden Studie aufgebaut wird, kommt es unweigerlich zu Überschneidungen mit Vorgängerarbeiten. Verwiesen sei u.a. auf Höhmann 2007, 2011, 2013a, 2013b, 2013c, 2014a, 2014b und 2016.

Literaturhinweise

- Atkins, S. and Rundell, Michael 2008. *The Oxford Guide to Practical Lexicography*. Oxford University Press.
- Baroni, Marco 2010. Corpora di italiano. In Raffaele Simone (ed.), *Enciclopedia dell'italiano*, vol 1. Roma: Istituto della Enciclopedia Italiana, 300–303 ([http://www.treccani.it/enciclopedia/corpora-di-italiano_\(Enciclopedia-dell'Italiano\)/](http://www.treccani.it/enciclopedia/corpora-di-italiano_(Enciclopedia-dell'Italiano)/)).

- Baroni, Marco / Bernardini, Silvia / Comastri, Federica / Piccioni, Lorenzo / Volpi, Alessandra / Aston, Guy / Mazzoleni, Marco 2004. Introducing the la Repubblica corpus. A large, annotated, TEI(XML)-compliant corpus of newspaper Italian. In Lino, Maria Teresa *et al.* (eds.), *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Paris: ELRA – European Language Resources Association, 1771–1774.
- Belica, Cyril. 2011. Semantische Nähe als Ähnlichkeit von Kookkurrenzprofilen. In Andrea Abel, Renata Zanin, Hrg., *Korpora in Lehre und Forschung*, Bozen-Bolzano University Press. Freie Universität Bozen-Bolzano, 155–178 (<http://pro.unibz.it/library/bupress/publications/fulltext/9788860460400.pdf>).
- Brandt, Wolfgang 1996. Handlungsobligationen und Handlungsoptionen. Modalverben und ihre verbalen Ersatzformen in der deutschen Gesetzessprache. In Hennig, Jörg / Meier, Jürgen (Hrg.), *Varietäten der deutschen Sprache. Festschrift für Dieter Möhn*. Frankfurt a.M. u.a., Peter Lang, 229–246.
- Brezina, Vaclav/ McEnery, Tony / Wattam, Stephen. 2015. Collocations in context. A new perspective on collocation networks. *International Journal of Corpus Linguistics* 20:2 (2015), 139–173.
- Busse, Dietrich. 2000. Textsorten des Bereichs Rechtswesen und Justiz. In Brinker, Klaus / Antos, Gerd / Heinemann, Wolfgang / Sager, Sven F. (Hrg.), 2000–2001. *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*. 2 Bde. Berlin u.a., de Gruyter, 658–675.
- Cap, Fabienne / Heid, Ulrich 2011. Distinguishing Specialised Discourse: The Example of Juridical Texts on Industrial Property Rights and Trademark Legislation. *Internal Relations Online Working Paper Series* 11/3, Stuttgart: Stuttgart University Press (http://www.uni-stuttgart.de/soz/ib/forschung/IRWorkingPapers/IROWP_Series_2011_3_Cap_Heid_Specialized_Discourse.pdf).
- Cheng, Winnie, Greaves, Chris and Warren, Martin. 2006. From n-gram to skipgram to con-gram. *International Journal of Corpus Linguistics* 11/4: 411–433.
- Cheng, Winnie, Greaves, Chris, Sinclair, John McH. and Warren, Martin. 2009. Uncovering the extent of the phraseological tendency: Towards a systematic analysis of congrams, *Applied Linguistics*, 30/2: 236–252.
- Didakowski, Jörg / Geyken, Alexander / Lemnitzer, Lothar. 2012. Automatic example sentence extraction for a contemporary German dictionary. In Ruth Vatvedt Fjeld and Julie Matilde Torjusen: *Proceedings of the 15th EURALEX International Congress*. 7-11 August 2012. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo, 343–349 (<http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202012/pp343-349%20Didakowski,%20Lemnitzer%20and%20Geyken.pdf>).
- Didakowski, Jörg, Geyken Alexander, 2013. From DWDS corpora to a German Word Profile – methodological problems and solutions. In *Network Strategies, Access Structures and Automatic Extraction of Lexicographical Information. 2nd Work Report of the Academic Network "Internet Lexicography"*. Mannheim: Institut für Deutsche Sprache. (OPAL - Online publizierte Arbeiten zur Linguistik X/2012), 43–52 (http://eins.dwds.de/static/website/publications/pdf/didakowski_geyken_internetlexikografie_2012_final.pdf).
- Dobrić Basanež, Katja. 2015. Investigating ‘congrams’ in the language of contracts and legal agreements. *Fachsprache: Internationale Zeitschrift für Fachsprachenforschung, -didaktik und Terminologie*, 2015 37/3-4, 176–192.

- Faaß, Gertrud / Heid, Ulrich / Prinsloo, Daan J. / Taljard, Elsabé. 2009. Part-of-Speech tagging of Northern Sotho: Disambiguating polysemous function words. In *Proceedings of the EACL 2009 Workshop on Language Technologies for the African Languages (AfLaT 2009)*. 31 March 2009, 38–45 (www.aflat.org/files/W09-0706.pdf).
- Felder, Ekkehard. 2012. Pragma-semiotische Textarbeit und der hermeneutische Nutzen von Korpusanalysen für die linguistische Mediendiskursanalyse. In Felder, Ekkehard / Müller, Marcus / Vogel, Friedemann (Hrsg.), *Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen*. Berlin / New York: de Gruyter, 115–174.
- Geyken, Alexander. 2011. Statistische Wortprofile zur schnellen Analyse der Syntagmatik in Textkorpora. In Andrea Abel, Renata Zanin (Hrsg.), *Korpora in Lehre und Forschung*. Bozen-Bolzano University Press, 115–137 (<http://pro.unibz.it/library/bupress/publications/fulltext/9788860460400.pdf>).
- Greaves, Chris. 2009. ConcGram 1.0. Amsterdam: John Benjamins.
- Gries, Stefan Th. 2014. Quantitative corpus approaches to linguistic analysis: seven or eight levels of resolution and the lessons they teach us. In Irma Taavitsainen, Merja Kytö, Claudia Claridge & Jeremy Smith (eds.), *Developments in English: expanding electronic evidence*. Cambridge: Cambridge University Press, 29–47.
- Gries, Stefan Th. 2015. Polysemy. In Ewa Dąbrowska & Dagmar S. Divjak (eds.), *Handbook of Cognitive Linguistics*. Berlin & Boston: De Gruyter Mouton, 472–490.
- Gries, Stefan Th. & Dagmar S. Divjak. 2009. Behavioral profiles: a corpus-based approach towards cognitive semantic analysis. In Vyvyan Evans & Stephanie S. Pourcel (eds.), *New directions in cognitive linguistics*, Amsterdam & Philadelphia: John Benjamins, 57–75.
- Gross, Gaston 2008. Les classes d'objets. *Lalies* 28, 111–165.
- Höhmman, Doris. 2007. Überlegungen zur qualitativen und quantitativen Untersuchung von Modalverben und ihren Konkurrenzformen in Rechtstexten. In Heller, Dorothee / Taino, Pierluigi (Hrsg.), *Italienisch-deutsche Studien zur fachlichen Kommunikation*, Frankfurt a.M., P. Lang, 25–39.
- Höhmman, Doris. 2011. *Lexikalische Konfigurationen. Korpusgestützte Mikrostudien zur Sprachlichkeit im deutschen und italienischen Verwaltungsrecht*, Peter Lang, Frankfurt a.M.
- Höhmman, Doris. 2013a. Korpusgestützte Wortfeldanalysen. Zum Auf- und Ausbau verwaltungsrechtlicher Sprachbestände. In *Die Rechtssprache in der internationalen Diskussion*. Hrsg. von Marina Wagnerová und Gerald G. Sander. Hamburg: Kovač, 43–57.
- Höhmman, Doris. 2013b. Zur Untersuchung erweiterter Nominalgruppen mit Hilfe von Concgrams. Eine sprachvergleichende Studie zum deutschen und italienischen Umweltrecht. In Cécile Desoutter, Dorothee Heller & Michele Sala (eds), *Korpora in specialized communication - Korpora in der Fachkommunikation - Les corpus dans la communication spécialisée*, CERLIS Series, vol.4, 267–296 (<http://dinamico.unibg.it/cerlis/page.aspx?p=276>)
- Höhmman, Doris. 2014a. Zur korpuslinguistischen Untersuchung der sprachlichen Ausgestaltung von Wissensrahmen. Eine Fallstudie anhand des thematischen Bausteins „Die Biennale in Venedig“. In Lombardi, Alessandra / Mor, Lucia / Roßbach, Nikola (eds.), *Reiseziel Italien. Moderne Konstruktionen kulturellen Wissens in Literatur – Sprache – Film*, Frankfurt: Peter Lang, 145–169.
- Höhmman, Doris. 2014b. Landschaftsdarstellungen in verschiedenen fach- bzw. domänenbezogenen Kommunikationszusammenhängen. Eine quantitative Annäherung. *mediAzioni* 16, 1–25 (<http://mediazioni.sitlec.unibo.it/index.php/no-16-2014/96-dossier-traduzione-specializzata-2014.html>).

- Höhmann, Doris. 2016. Il “corpus del rinascimento”. Stato dell’arte e prospettive. In Mari Pietrogiovanna (ed.), *Uno sguardo verso nord*. Scritti in onore di Caterina Virdis Limentani. Padova: *Il Poligrafo*, 217–231.
- Höhmann, Doris / Spissu, Maria Vittoria. 2013. Il “corpus del Rinascimento”. Strumenti per la comunicazione in ambito specialistico tra storia dell’arte e turismo culturale. In D. Höhmann (Hrsg.), *Tourismuskommunikation. Im Spannungsfeld von Sprach- und Kulturkontakt*, Peter Lang, Frankfurt a.M., 349–381.
- Holl, Daniel 2001. Was ist modal an Modalen Infinitiven? In Müller, Reimar / Reis, Marga (Hrsg.), *Modalität und Modalverben im Deutschen*. Linguistische Berichte. Sonderheft. Hamburg: Helmut Buske Verlag, 217–238.
- Jakubíček, Miloš / Kilgarriff, Adam / Kovář, Vojtěch / Rychlý, Pavel / Suchomel, Vít. 2013. The TenTen Corpus Family. 7th International Corpus Linguistics Conference CL 2013, Lancaster. Lancaster University, 125–127 (https://www.sketchengine.co.uk/wp-content/uploads/The_TenTen_Corpus_2013.pdf).
- Keßler, Christine 2008. Zur semantischen Beschreibung von Zahlwörtern im System und im Text. In Pohl, Inge (Hrsg.), *Semantik und Pragmatik – Schnittstellen*. Frankfurt a.M. [u.a.].
- Kilgarriff, Adam, Rychlý, Pavel, Smrz, Pavel, Tugwell, David 2004. ‘The Sketch Engine’. In Williams, Geoffrey / Vessier Sandra (eds.), *Proceedings of the 11th EURALEX International Congress*. 6–10 July 2004, *Lorient, France*, 105–116 (<http://euralex.org/publications/the-sketch-engine/>).
- Kilgarriff, Adam / Husák, Miloš, / McAdam, Katy, Rundell, Michael, Rychlý, Pavel 2008. Gdex: Automatically finding good dictionary examples in a corpus. In: Elisenda Bernal & Janet DeCesaris (eds.), *Proceedings of the XIII EURALEX International Congress*. 15–19 July 2008. Barcelona: Universitat Pompeu Fabra, 425–432 (<http://euralex.org/category/publications/euralex-2008/>).
- Kilgarriff, Adam / Rychlý, Pavel / Kovář, Vojtěch / Baisa, Vít. 2012. Finding Multiwords of More Than Two Words. In Ruth Vatvedt Fjeld and Julie Matilde Torjusen (eds.), *Proceedings of the 15th EURALEX International Congress*. 7-11 August 2012. Oslo: *Department of Linguistics and Scandinavian Studies*, University of Oslo, 693–700 (<http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202012/pp693-700%20Kilgarriff,%20Rychly,%20Kovar%20and%20Baisa.pdf>).
- Kübler, Sandra. 2010. Baumbanken. In Carstensen, Kai-Uwe / Ebert, Christian / Ebert, Cornelia / Jekat, Susanne / Klabunde, Ralf / Langer, Hagen (Hrsg.) 2010, *Computerlinguistik und Sprachtechnologie. Eine Einführung*. Heidelberg, Spektrum Akademischer Verlag, 492–503.
- Kupietz, Marc / Belica, Cyril / Keibel, Holger / Witt, Andreas 2010. The German Reference Corpus DeReKo: A primordial sample for linguistic research. In Calzolari, Nicoletta et al. (eds.), *Proceedings of the 7th conference on International Language Resources and Evaluation* (LREC 2010). Valletta, Malta: European Language Resources Association (ELRA), 1848–1854 (http://www.lrec-conf.org/proceedings/lrec2010/pdf/414_Paper.pdf).
- Kupietz, Marc/Lüngen, Harald. 2014. Recent Developments in DeReKo. In Calzolari, Nicoletta et al. (eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (LREC’14). Reykjavik: ELRA, 2378–2385 (http://www.lrec-conf.org/proceedings/lrec2014/pdf/842_Paper.pdf).
- Lombardi Alessandra 2013. ‘Usuelle Wortverbindungen und wiederkehrende Kommunikations-handlungen in touristischen Texten. Das Projekt TouriTerm’. In Höhmann, D. (Hrsg.), *Tourismuskommunikation. Im Spannungsfeld von Sprach- und Kulturkontakt*. Frankfurt a.M. u.a.: Peter Lang, 323–347.
- Lyons, John 1968: *Introduction to Theoretical Linguistics*. London: Cambridge University Press.

- Panunzi, Alessandro / Cresti, Emanuela / Gregori, Lorenzo 2014. RIDIRE. Corpus and tools for the acquisition of Italian L2. In Abel Andrea / Vettori, Chiara / Ralli, Natascia (eds.), *Proceedings of the XVI EURALEX International Congress, 15-19 July 2014, Bolzano-Bozen, 447-462* (http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202014/euralex_2014_033_p_447.pdf).
- Prinsloo, Danie J. / Heid, Ulrich / Bothma, Theo / Faaß, Gertrud 2012. Devices for Information Presentation in Electronic Dictionaries. In *Lexikos* 22, 290-320 ([http://www.repository.up.ac.za/dspace/bitstream/handle/2263/21127/Prinsloo_Devices\(2012\).pdf?sequence=1](http://www.repository.up.ac.za/dspace/bitstream/handle/2263/21127/Prinsloo_Devices(2012).pdf?sequence=1)).
- Roelcke, Thorsten 1989. *Die Terminologie der Erkenntnisvermögen. Wörterbuch und lexikosemantische Untersuchung zu Kants „Kritik der reinen Vernunft“*. Tübingen: Niemeyer.
- Roelcke, Thorsten 2004. Stabilität statt Flexibilität? Kritische Anmerkungen zu den semantischen Grundlagen der modernen Terminologielehre. In Pohl, Inge/ Konerding, Klaus-Peter (Hrsg.), *Stabilität und Flexibilität in der Semantik. Strukturelle, kognitive, pragmatische und historische Perspektiven*. Frankfurt/M.: Peter Lang, 2004, 137-150.
- Roelcke, Thorsten 2010³. *Fachsprachen*. Berlin: Schmidt.
- Rossini Favretti, R. 2000. Progettazione e costruzione di un corpus di italiano scritto: CORIS/ CODIS. In R. Rossini Favretti (ed.), *Linguistica e informatica. Multimedialità, corpora e percorsi di apprendimento*, Roma, Bulzoni, 39-56.
- Schnörch, Ulrich / Storjohann, Petra 2012. Ein Korpus als Garant zuverlässiger lexikografischer Informationen? Eine vergleichende Stichprobenuntersuchung. In Ruth Vatvedt Fjeld and Julie Matilde Torjusen (eds.), *Proceedings of the 15th EURALEX International Congress. 7-11 August 2012. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo, 310-322* (<http://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202012/pp310-322%20Schnorch%20and%20Storjohann.pdf>).
- Soffritti, Marcello 2000. Categorie pragmatostuali, precisione e polisemia in codici tedeschi e italiani: i verbi modali. In Schena, Leandro / Snel Trampus, Rita D. (a c. di), *Traduttori e giuristi a confronto*, Bologna: CLUEB, 37-53.
- Storjohann, Petra 2003. *Computergestützte Lesartendisambiguierung*. In *Deutsche Sprache* 1/03, 3-28.
- Storrer, Angelika 2006. Funktionen von Nominalisierungsverbgefügen im Text. Eine korpusbasierte Fallstudie. In Prost, Kristel / Winkler, Edeltraud (Hrsg.), *Von der Intentionalität zur Bedeutung konventionalisierter Zeichen. Festschrift für Gisela Harras zum 65. Geburtstag*. Tübingen: Narr, 147-178.
- Stubbs, Michael 2002. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Teubert, Wolfgang. 2005. My Version of Corpus Linguistics. In *International Journal of Corpus Linguistics* 2005, 10/1, 1-13.
- Weinrich, Harald. 2003. *Textgrammatik der deutschen Sprache*. Hildesheim u.a., Olms.
- Winkler, Edeltraud. 2009. Überlegungen zu Artefaktbezeichnungen im Deutschen. In *Deutsche Sprache* 37/1, 33-47.