

# Benutzerdokumentation

Technical Report IDS-KL-2007-02  
zum Produkt

## Korpusbasierte Wortgrundformenliste DEREWO v-30000g-2007-12-31-0.1

Institut für Deutsche Sprache, Mannheim  
Dezember 2007

### 1 Wie dieses Dokument zu lesen ist

Dieses Dokument versucht, die Dokumentation der aktuellen Version der Wortgrundformenliste mit Hintergrundinformationen zu verbinden. Die konkreten Entscheidungen bei der aktuellen Version zu verschiedenen Fragestellungen sind in Abschnitt 6.4 zusammengefasst. Wenn Sie die Liste im Kontext Ihrer Arbeit verwenden wollen, sollten Sie auf jeden Fall die Lizenzbestimmungen (Kap. 9) und den genannten Abschnitt berücksichtigen. Grundsätzlich empfehlen wir jedoch, das gesamte Dokument zu lesen.

### 2 Vorwort

Das Institut für Deutsche Sprache erreichen immer wieder Anfragen nach Häufigkeitsangaben für Wörter, insbesondere werden Listen gewünscht der „N häufigsten deutschen Wörter“.

Auf Rückfragen, für welche Fragestellung die Angaben genutzt werden sollen, welche Sprachgemeinschaft in welchem Zeitraum gemeint ist, was gezählt und worüber ggf. kumuliert werden soll, welche Einheiten außer acht gelassen werden können oder sollen, ob beim Bestimmen der ersten N noch besondere Aspekte zu beachten sind usw. zeigt sich oft erheblicher Klärungsbedarf.

So überflüssig und detailverliebt diese Rückfragen der Sprachwissenschaftler auf eine derart einfache Anfrage auch erscheinen mögen, so folgenreich sind die Antworten für die „korrekte“ Beantwortung der ursprünglichen Anfrage. Denn und das sollte beim Lesen dieser Dokumentation deutlich werden eine einzige, die „Top-N-Liste der deutschen Wörter“ gibt es einfach nicht. Die vielen einzelnen, für je eine bestimmte Sprachbetrachtungsperspektive angemessenen Ranglisten unterscheiden sich sowohl in ihrer Zusammensetzung wie auch in der Reihenfolge ihrer Einträge beträchtlich.

Mit der Veröffentlichung der hier dokumentierten Wortliste DEREWO bemüht sich das Institut für Deutsche Sprache einen Kompromiss zu finden zwischen der faszinierenden Vielfalt unserer sprachlichen Realität und dem berechtigten Wunsch nach ihrer möglichst kompakten, wenn auch teilweise vereinfachenden Beschreibung. Die DEREWO-Wortliste soll nicht den allgemein formulierten Wunsch aus dem ersten Absatz erfüllen, sondern geht auf ihn aus einem einzigen der vielen denkbaren Blickwinkel in wesentlichen Aspekten nur ein. Diese Dokumentation soll Ihnen dabei helfen die von uns gewählte Betrachtungsperspektive und die daraus resultierenden Folgen für die Interpretation und Handhabung der Liste zu erkennen.

Lesen Sie dieses Dokument bitte sorgfältig durch, um unnötige Missverständnisse oder Irrtümer, terminologische Verwechslungen oder methodische Fehlentscheidungen bei der Nutzung der DEREWO -Liste zu vermeiden.

Die DEREWO -Wortliste ist keine „Freeware“, sondern ein Werk im Sinne des Urheberrechts. Vergewissern Sie sich vor einer Nutzung, dass Sie die Nutzungsbedingungen gelesen und verstanden haben. Eine nicht erlaubte Nutzungsart kann Rechtsfolgen nach sich ziehen.

### 3 Download

Die Originale der DEREWO-Wortlisten können unter <http://www.ids-mannheim.de/kl/derewo/> zusammen mit der Dokumentation in der jeweils aktuellen Version abgerufen werden.

### 4 Was ist DEREWO?

DEREWO ist die Lemmastrecke eines fiktiven Wörterbuchs des öffentlichen Schriftsprachgebrauchs der letzten 30 Jahre im Umfang von 30.000 Lemmata zusammengestellt auf der Grundlage herkömmlicher lexikographischer Kriterien mit besonderer Berücksichtigung der Gebrauchshäufigkeit in Form von Korpusfrequenz.

### 5 Wie wurde DEREWO erstellt?

DEREWO basiert auf dem DEUTSCHEN REFERENZKORPUS DEREKO. Die DEREWO-Wortliste wurde in einer Kombination von automatischen, semi-automatischen und manuellen Verfahren erstellt, je nachdem, welche Vorgehensweise zur Lösung welcher partiellen Problemstellung sinnvoll bzw. erforderlich war.

### 6 Methodik im Einzelnen

#### 6.1 Korpusbasiertheit

Die dauerhafte Sicherung einer empirischen Grundlage für die germanistisch-sprachwissenschaftliche Forschung ist eine der zentralen Aufgaben des IDS. Zu diesem Zweck unterhält das Institut seit 1964 eine umfangreiche elektronische Stichprobe deutschsprachiger Texte aus Gegenwart und jüngerer Vergangenheit: das so genannte DEUTSCHE REFERENZKORPUS (DEREKO).

##### Das DEUTSCHE REFERENZKORPUS

- bildet mit über zwei Milliarden Wörtern die weltweit größte Sammlung elektronischer Korpora mit geschriebenen deutschsprachigen Texten aus Gegenwart und neuerer Vergangenheit
- enthält belletristische, wissenschaftliche und populärwissenschaftliche Texte, eine große Zahl von Zeitungstexten sowie eine breite Palette weiterer Textarten
- enthält ausschließlich urheberrechtlich abgesichertes Material
- ist mit vielen Metainformationen ausgezeichnet
- wird im Hinblick auf Umfang, Variabilität und Qualität kontinuierlich weiterentwickelt
- ist zu einem großen Teil kostenlos über die Recherchesoftware COSMAS II zugänglich
- erlaubt in der Nutzungsphase die Komposition virtueller Korpora, die repräsentativ oder auf spezielle Aufgabenstellungen zugeschnitten sind

Wir versuchen, mit unseren Korpora den Sprachgebrauch abzubilden. Die Aussagen, die wir aus der Auswertung unserer Korpora ableiten, lassen sich aber aufgrund der Datenzusammensetzung und anderer Faktoren nur bis zu einem gewissen Grad auf „den Sprachgebrauch“ verallgemeinern. Allein aber über die Größe des Datenbestandes können wir viele Effekte ausgleichen. Die Frequenz der sprachlichen Einheiten ist für uns ein wichtiges (wenn nicht sogar: das wichtigste) Kriterium für ihre Relevanz.

Dadurch, dass der Datenbestand viele Zeitungstexte enthält, ist er besonders gut geeignet, den

„öffentlichen Schriftsprachgebrauch“ abzubilden. Bei der Interpretation der Angaben ist jedoch im Hinterkopf zu behalten, dass der Datenbestand sich über die letzten 30 Jahre erstreckt, dadurch viele Texte in verschiedenen Schreibweisen der verschiedenen Rechtschreibreform-(akzeptanz-)phasen enthält, sowie eine Mischung darstellt von Texten aus der Bundesrepublik Deutschland, der Schweiz und aus Österreich. Frequenzangaben stellen somit immer eine Kumulation von Frequenzen in Texten mit diesen verschiedenen Eigenschaften dar. Durch den langen Zeitraum wirken z.B. die Häufigkeiten in älteren Texten z.B. bei älteren Realien (wie die Währungsangabe *Mark*) noch nach.

Was unter den sprachlichen Einheiten verstanden werden kann, variiert so sehr, wie der Begriff „Wort“ unklar gehandhabt wird.

## 6.2 Wortformenlisten

Im einfachsten Fall kann man unter „Wort“ eine konkrete Wortform verstehen, d.h. die Realisierung einer alphabetischen (oder alphanummerischen?) Zeichenkette in einem (hier: geschriebenen) Text. Dazu bedarf es lediglich festzulegen, welche Bestandteile zu einer Wortform gehören, welche Zeichen Wortformen trennen und wie Groß-/Kleinschreibung und evtl. Zweifelsfälle zu handhaben sind.

Für unser konkretes Beispiel gehen wir von folgenden Annahmen aus bzw. legen wir folgende Vereinbarungen fest: Für die deutsche Sprache sind die Bestandteile die alphabetischen Zeichen a-z, A-Z inkl. der Umlaute ä, ö, ü, Ä, Ö und Ü sowie das ß (demnächst auch als Majuskel). Die Trenner sind Satzzeichen (,;:?!), Leerzeichen und Zeilenumbrüche (außer bei Worttrennung am Zeilenende). Unklare Fälle sind z.B. der Bindestrich (-), der Punkt (. Satzende oder *Dr./bzw.*) und der Trennstrich (beim *Zeilen-¶umbruch*). Trennstriche beim Zeilenumbruch werden normalerweise aufgelöst (d.h. die Bestandteile auf den verschiedenen Zeilen ohne Trennstrich zusammengezogen, Spezialfall *kk* wird wieder zu *ck*: *Zuk-kerzu Zucker*), der Punkt wird außer in einer endlichen Menge von aufgezählten Ausnahmefällen als Satzende und somit als Trennzeichen interpretiert. Der Bindestrich ist gerade in Zeiten von Internetadressen und den verschiedenen Phasen der Rechtschreibreform sehr schwierig einheitlich zu handhaben. Er wird in unserem Fall nicht als Wortbestandteil betrachtet. Zwischen Groß- und Kleinschreibung wird auf Wortformenebene konsequent unterschieden.

Nachdem für die genannten Bereiche Konventionen vereinbart wurden, ist es leicht, den Text in die relevanten Zeichenketten zu zerlegen (zu tokenisieren), alle Vorkommen der Zeichenketten (der Tokens) durchzugehen und die Häufigkeit aller Zeichenketten mit demselben Erscheinungsbild zu kumulieren. Damit erhält man die Häufigkeit der Types. Der Type *alle* hat im vorletzten Satz z.B. die Häufigkeit 2, da zwei Tokens dieser Form im Satz gezählt werden.<sup>1</sup>

## 6.3 Grundformenlisten

In vielen Fällen sind Wortformenlisten nicht adäquat, z.B. für eine Stichwort(kandidaten)liste für Lexikographen oder ähnliche Untersuchungen. In solchen Fällen ist eine Liste von Grund- oder Nennformen geeigneter.

In einem Flektions- bzw. Wortbildungsparadigma eines „Wortes“ steckt viel Redundanz: Man kann alle wichtigen Informationen zu einer (Nenn-)Form angeben, die Angaben zu allen Wortformen des Paradigmas wären identisch, die Formen unterschieden sich – bis zu einem gewissen Grad – nur durch die unterschiedliche Realisierung aufgrund der Morphologie (o.ä.).

---

1 Für Wortformenlisten richten Sie Ihre Anfrage bitte an das Projekt COSMAS II ([cosmas2@ids-mannheim.de](mailto:cosmas2@ids-mannheim.de)).

Die Frequenz dieser Nennform, des sogenannten Lemmas, ist die kumulierte Häufigkeit aller Wortformen, die zu dem Paradigma beitragen. Für psychologische Untersuchungen wird ein Begriff der Geläufigkeit häufig in Verbindung zur Frequenz gesetzt. Dabei unreflektiert über Wortformen zu argumentieren ist bedenklich: In manchen Paradigmen gibt es z.B. viele homonyme Formen, die dadurch einen Anstieg der Häufigkeit bedingen (s. vor allem auch Formen vom Typ *be-X-t*: *beklagt*, *behauptet*; gleichzeitig 3.PersSingPräsIndAkt, PartPerf und somit auch adjektivisch verwendbar, s.u.); diese werden im Vergleich zu Wörtern mit vielfältigen Formen im Paradigma quasi überbewertet. Unstrittig ist wohl, dass, aber unklar ist noch, inwieweit sich morphologisch (o.ä.) Verwandtes gegenseitig bei „Geläufigkeit“ beeinflusst. Für verschiedene Fragestellungen wäre es deshalb wünschenswert, einen Begriff einer „Grundform“ zu haben, auf den man sich beziehen kann. *Auch wenn wir diesen Begriff nicht vollständig klären bzw. definieren können, wollen wir uns wenigstens in die Richtung bewegen und uns ihm annähern.*

Bei der Definition des Begriffs „Grundform“ wirken zum einen natürlich die o.g. Entscheidungen nach (insbesondere Groß-/Kleinschreibung und Trennzeichen) bzw. müssen hier erneut ausgehandelt werden; zum anderen ergeben sich viele weitere Fragen. Manche dieser Fragen lassen sich nur über die Kompetenz eines Sprechers bzw. durch eine tiefe Analyse der jeweiligen Fälle beantworten; beides ist in unserem Szenario nicht möglich, da unser Datenbestand so umfangreich ist, dass diese Aufgaben nur mit vollautomatischen Verfahren in einem vertretbarem Zeitrahmen bearbeitet werden können. Da wir an der Authentizität der Daten festhalten und somit keine vermeintlichen Fehler korrigieren, sind vollautomatische Verfahren für eine derartige Analyse utopisch (bestehende Ansätze greifen nicht und wir können aus verschiedenen Gründen dieses Anliegen nicht verfolgen). In manchen Fällen versuchen wir, kleine Testmengen von Hand bzw. halbautomatisch auszuwerten. Diese Ergebnisse werden, soweit es geht, auf die Gesamtheit extrapoliert.

Im Gegensatz zu einem Stamm bzw. zu einer Stammform, die man als Rumpf definiert ohne Flexionsendungen oder sonstige Suffixe oder auch Präfixe, meint man mit Grundform gemeinhin eine Form aus dem Paradigma, die auch realisiert werden kann (was beim Stamm nicht zwingend der Fall ist). Bei Substantiven ist dies standardmäßig der Nominativ Singular, bei Verben der Infinitiv, bei Adjektiven der Positiv. Bereits dabei ergeben sich Probleme: Sind adjektivisch gebrauchte Partizipien auf den Positiv oder auf den Infinitiv des zugrunde liegenden Verbs zurückzuführen? Präverbfügungen (früher auch „präfigierte Verben“) manifestieren sich in Texten z.T. als kontinuierliche, z.T. aber auch als diskontinuierliche Elemente. Letztere sollten sicher derselben Grundform (dem Infinitiv der Präverbfügung) zugeordnet werden wie die ersteren, stellen aber ein großes Problem für die automatische Erkennung dar.

Von der Angabe einer Grundform wird nur an wenigen Stellen abgewichen, wenn wir der Meinung waren, dass aus der Form nicht ersichtlich ist, welche weiteren Formen sich dahinter verbergen. Als Ersatz geben wir in den Fällen eine mit Bindestrich gekennzeichnete quasi-Stammform an, z.B. *d-* für alle Formen der Artikel *der*, *die*, *das*.

### 6.3.1 Groß-/Kleinschreibung

Bei der Tokenisierung wurde zwischen Groß- und Kleinschreibung unterschieden. Für die Bestimmung einer Grundform ist die Unterscheidung aber ein wesentlich unzuverlässigerer Indikator als gemeinhin angenommen werden könnte: Am Satzanfang, in Überschriften, bei Titeln, festen Wendungen (*Tag der Offenen Tür*, vgl. Rechtschreibreform), stilistischen Spielereien, bei Zitaten und Hervorhebungen wird durchaus von der üblichen Variante abgewichen. Dabei wird nicht nur groß- statt kleingeschrieben. Das Internet macht sich auch

hier bemerkbar: Viele Varianten fließen auch durch die Bezugnahme auf WWW- und Email-Quellen und die Angabe der entsprechenden Adressen (die viele Wörter der deutschen Sprache in Kleinschreibung enthalten) in die Sprache ein. Mit dem gleichen Effekt, aber ebenfalls für uns nicht ohne weiteres oberflächlich erkennbar, werden Konversionen in der Sprache gebraucht (Nullderivationen: substantivierte Verben und Adjektive, Denominalisierungen), bei denen evtl. darüber nachgedacht werden müsste, ob für diese eigene Grundformen anzusetzen sind. Zunächst sehen wir aber davon ab und übernehmen bis auf Einzelfälle den Vorschlag des automatischen Lemmatisierers: Verbindungen von Substantiv und Partizip führt der Lemmatisierer z.T. auf das dem Partizip zugrundeliegende Verb zurück. Dies haben wir in den erkannten Fällen durch das Partizip (*vertrauensbilden* zu *vertrauensbildend*, in einem Fall durch ein Substantiv ersetzt (*wirtschaftstreiben* zu *Wirtschaftstreibender*). Großgeschriebene Präpositionen (*Für* sehr wahrscheinlich häufiger am Satzanfang als in *das Für [und Wider]*) haben wir in den erkannten Einzelfällen der kleingeschriebenen Variante zugeordnet.

Mittel- und langfristig wäre zu überlegen, die quantitativen Verteilungen innerhalb der vermeintlichen Paradigmen mit zu der Entscheidung heranzuziehen, wie viele bzw. welche der Formen als Grundformen anzusetzen sind. Interessant erscheint auch die Idee eines kontrastiven Vergleichs der Kookkurrenzprofile der verschiedenen Wortformen eines Lemmas.

### 6.3.2 Trennzeichen/Bindestrich

Der Trennstrich ist bei der Tokenisierung aufgelöst worden. Der Bindestrich wird in unserem konkreten Fall nicht als Bestandteil einer Wortform gedeutet.

### 6.3.3 Neubildungen/Neologismen

Unser Bestreben ist es, mit unseren Korpora ein Abbild des Sprachgebrauchs und natürlich auch von dessen Wandel nachzuzeichnen. Der Begriff einer „Grundform“ stößt dabei an gewisse Grenzen, weil wir damit vorwegnehmen müssen, für welche verschiedenen Formen eine vom Sprecher/Schreiber intendierte gemeinsame Grundform anzunehmen ist. Manche Paradigmen von neuen Wortformen lassen sich leicht systematisch erschließen. Durch die Rechtschreibreform sind neue Schreibweisen eingeführt worden, die Flexion entspricht aber weitestgehend den alten Regularitäten. Für unseren Fall haben wir versuchsweise die vom Lemmatisierer nicht erkannten Formen auf partielle Übereinstimmungen mit Adjektivparadigmen überprüft. Die manuelle Auswertung hat z.T. neue Adjektivschreibweisen (z.B. *rau*, *aufwändig*) aufgedeckt, z.T. auch Hinweise auf Adjektive, die – entgegen bisheriger Postulate – (mittlerweile) gesteigert verwendet werden (*politischste*, *universellste*). Die erkannten Neuerungen wurden ggf. als neue Grundform bzw. mit einer zusätzlichen Kumulierung der Häufigkeiten der weiteren Wortformen eingearbeitet.

Bei Verben wurde analog lediglich in Einzelfällen, die beim Wörterbuchabgleich aufgefallen sind (*platzieren*), eine Grundform für ein Verbparadigma mit kumulierten Häufigkeiten eingeführt. Da Neubildungen bisher nicht systematisch erfasst sind, verbirgt sich darin die Gefahr, dass sie zu Unrecht nicht für unsere Auswahl berücksichtigt wurden, obwohl die kumulierte Frequenz ihres Paradigmas die Aufnahme gerechtfertigt hätte.

### 6.3.4 Fremdwörter, Anglizismen

Auch wenn Fremdwörter, insbesondere Anglizismen sehr häufig in der deutschen Sprache gebraucht werden, ist doch nicht unumstritten, inwieweit sie zu einem Kernbestand der deutschen Sprache gehören. Für die aktuelle Version haben wir uns dazu entschieden, in einem Vergleichswörterbuch belegte Fremdwörter zu akzeptieren. Die nicht belegten Fremdwörter,

insbesondere die Anglizismen wurden akzeptiert, sofern es sich um Inhaltswörter handelte, die in den vermuteten allgemeinen Sprachgebrauch eingeflossen sind. Funktionswörter und unseres Erachtens genretypische Bezeichnungen wurden herausgefiltert.

Durch unsere Vorauswahl aus der Wortformenliste durch das deutsche Alphabet bleiben authentische Schreibweisen von Fremdwörtern aus Sprachen, die abweichende Alphabete benutzen (z.B. das Französische mit diakritischen *e*: *é, è, ê*), im Gegensatz zu Fremdwörtern aus Sprachen, die dasselbe Alphabet benutzen (wie z.B. das Englische), unberücksichtigt.

### 6.3.5 diskontinuierliche Konstituenten, Präverbfügungen (abtrennbare/abgetrennte Präfixe)

*„Der Mensch stammt vom Affen ab. Es stimmt, dass der Mensch vom Affen abstammt. Der Mensch – vom Affen abstammend – ...“*

Das Ergebnis des Tokenisierens unterscheidet zwischen den Wortformen *stammt*, *ab*, *abstammt* und *abstammend*, auch wenn im ersten Satz lediglich eine diskontinuierliche Realisierung derselben Grundform *abstammen* (wie in den anderen beiden Sätzen) zu verzeichnen ist. Uns steht keine Möglichkeit zur Verfügung, diesen Zusammenhang an der Oberfläche exhaustiv zu erkennen. Als Ersatz bedienen wir uns eines Moduls *adjVerbFreq*: Mit Hilfe einer Testmenge wird ermittelt, wie oft die oberflächlich erkennbaren Konstituenten, d.h. die Präpositionen (die zumeist formgleich zu Präfixen sind), die rudimentären Verben ohne Präfix und die präfigierten Formen gezählt wurden; eine zusätzliche Informationsquelle gibt uns (für die Testmenge) einen Hinweis darauf, in wie vielen Fällen die vermeintlichen Präfixe tatsächlich als echte Präfixe zu deuten sind und zu welcher Fügung sie sich dann mit welchem rudimentären Verb verbinden. Die Frequenz der Präpositionen wird dann um die Frequenz der (echten) Präfixe reduziert, die Frequenz der Präverbfügungen wird entsprechend erhöht und die Frequenz der rudimentären Verben (quasi die „Nettoverben“) analog verringert.

Bsp.

<i>abstammen</i>	wird um geschätzte Vorkommen von <i>stammen</i> mit abgetrenntem <i>ab</i> erhöht
<i>ab</i>	wird u.a. um geschätzte Vorkommen von <i>stammen</i> und abgetrenntem <i>ab</i> verringert (sowie um evtl. weitere Vorkommen anderer Verben mit dem Präfix <i>ab</i> )
<i>stammen</i>	wird u.a. um geschätzten Vorkommen von <i>stammen</i> und abgetrenntem <i>ab</i> verringert (sowie um evtl. weitere Vorkommen von <i>stammen</i> mit anderem abgetrenntem Präfix)

Die Vorschläge des Moduls wurden von Hand auf Plausibilität überprüft, insbesondere diejenigen Kandidaten, die sich am stärksten in ihrer Häufigkeitsklasse (s.u.) verändert haben, und diejenigen, bei denen die Größenordnung derart adjustiert wurde, dass sie dadurch erst ein Kandidat für die Zielmenge geworden sind. In den Zweifelsfällen wurde die Adjustierung in der Richtung zwar beibehalten, aber auf eine Verdoppelung bzw. Halbierung (Änderung um eine Häufigkeitsklasse) gedämpft.

Andere Formen diskontinuierlicher Konstruktionen, wie z.B. *Im-* und *Export*, *auf-* und *abbauen*, werden vorläufig nicht explizit behandelt. Hier könnten wir uns ähnlicher Techniken bedienen, evtl. mit der zusätzlichen (noch zu überprüfenden) Annahme, dass die Häufigkeitsklassen der vervollständigten Formen nicht allzu weit auseinander liegen dürften.

### 6.3.6 Eigennamen

Wenn wir annehmen, dass wir die ersten 10.000, 20.000 oder 30.000 Einträge einer Grundformenliste bestimmen können, gehen wir auch davon aus, dass wir die Listen nach

Gewichtigkeit sortieren können. Unser wichtigstes Kriterium dafür ist die Frequenz. Unter den ersten ca. 37.000 Einträgen finden sich allerdings ca. 5.000 Formen, die man im weitesten Sinne als Eigennamen auffassen kann. Auch unabhängig von der Zusammensetzung unserer Datengrundlage (viele Zeitungen, in denen viel über Personen aus Politik, Wirtschaft, Kultur und Sport berichtet wird), erscheint fragwürdig, ob unsere Auswahl auch mit einer gedämpften Anzahl an Eigennamen den Bestand dessen wiedergibt, was wir als Kern einer Sprache betrachten würden. Um eine Sprache zu erlernen, braucht man dieses große Repertoire an Eigennamen sicher nicht; um eine Zeitung lesen zu können evtl. schon, nur ist in dem Kontext auch selbstverständlich, dass dazu andere Wissensquellen herangezogen werden müssen. Unsere Idee ist deshalb, Eigennamen möglichst aus unserer Wortgrundformenliste herauszufiltern, und ggf. eine separate Liste der verzeichneten Eigennamen ergänzend dazu herauszugeben. Den letzten Punkt werden wir aber erst später realisieren.

Die Erkennung der Eigennamen ist verhältnismäßig vielleicht sogar das „kleinere Problem“. Ähnlich wie bei den diskontinuierlichen Konstituenten können wir uns einer weiteren Informationsquelle bedienen. Alternativ können wir mit einem Verfahren uns nach und nach zu einer präziseren Eigennamenmenge hochschaukeln: Beginnend mit Wortformen, die relativ eindeutig meistens vor Eigennamen stehen (*Herr, Frau, Prof., Präsident, Kanzler, Vorsitzende* usw.) lassen sich die danach folgenden Wortformen zunächst quantitativ und dann qualitativ auswerten. Mit ihrer Hilfe kann man in die andere Richtung zurück schauen, welche Wortformen typischerweise davor stehen. Dieser Prozess lässt sich ggf. mehrmals wiederholen, bis sich die Menge stabilisiert. Ein zweites, und sehr wahrscheinlich das größere Problem ist allerdings die Bewertung, wie oft tatsächlich eine Wortform als Eigenname verwendet wird (*Kohl*/vs. *Kohl*) und auf wie viele verschiedene Instanzen damit referiert werden kann. Denn erst daraus lässt sich ein Zusammenhang zwischen Frequenz und Relevanz ableiten. Um diese Frage zu beantworten, könnte man (a) kleine Testmengen von Hand auswerten und dann extrapolieren oder (b) sich das Kookkurrenzverhalten der Wortformen anschauen, da sich die Heterogenität (Referenz auf viele Personen bzw. auf Person und Objekt) in einer Partitionierung des Kontextverhaltens zeigen sollte.

Für diese aktuelle Version haben wir es derart gehandhabt, dass alle nicht in einem zum Vergleich herangezogenen Wörterbuch verzeichneten Eigennamen bei der Festlegung der Zielmenge herausgefiltert wurden.

### 6.3.7 adjektivisch gebrauchte Partizipien

Bei vielen adjektivisch gebrauchten Partizipien ist noch vollkommen transparent, von welchem Verb sie abgeleitet wurden, etwa *das entspannende Bad*. Obwohl flektiert eindeutig als Adjektiv erkennbar, spüren wir beim Lesen noch die große Nähe zu dem Verb *entspannen*. Anders bei gewissermaßen verblassten Bildungen wie *das spannende Buch*. Es ist durchaus nachvollziehbar, wenn wir intuitiv einmal das Adjektiv *spannend* (genaugenommen das Partizip Präsens des Verbs *spannen*), das andere Mal aber das Verb *entspannen* als Grundform ansetzen würden. In anderen Fällen ist das Partizip (Perfekt) homonym zu anderen Formen des Paradigmas, teilweise zur dritten Person Singular (*er (hat) behauptet*) oder zur dritten Person Plural/Infinitiv (*wir verlaufen uns/ wir haben uns verlaufen*), so dass unklar ist, wenn wir diese als Grundformen anzusetzen wie welche Wortformen auf diese verteilt werden können. Der automatische Lemmatisierer ist bestrebt, sich in dem einen oder anderen Fall jeweils begründet für eine sinnvolle Alternative zu entscheiden. Im Zweifelsfall hat es sich als praktikabler erwiesen, zunächst eher eine Wortform zuviel anzunehmen, die über einen Wörterbuchlemmastreckenabgleich und manuelle Kontrolle evtl. als unzutreffend erkannt und dem allgemeineren Paradigma zugeordnet werden konnte. Für unsere aktuelle Version sind wir größtenteils den

Vorschlägen des Lemmatisierers gefolgt und haben uns nur bei der Kontrolle je nach Einzelfall für die eine oder Variante entschieden.

Auch hier wäre es lohnenswert, sich das Kookkurrenzverhalten des Wortformen anzuschauen, da sich die Heterogenität (als Rechtfertigung für die Ansetzung mehrerer Grundformen) in einer Partitionierung des Kontextverhaltens zeigen sollte.

### 6.3.8 Regionalismen

Aufgrund der Zusammensetzung der Datengrundlage kann es sein, dass Wortformen aus der Region um Mannheim herum überrepräsentiert sind. Dazu wurde eine Wortformenliste aus dem „Mannheimer Morgen“-Subkorpus erstellt. Alle Einträge wurden statistisch im Hinblick auf das gesamte Korpus bewertet, inwieweit sie auffällig oft/zu stark im MM-Korpus vertreten sind. Diese Kandidaten für Regionalismen wurden von Hand auf Plausibilität überprüft. Bei den bestätigten Regionalismen wurde die Wortformhäufigkeit auf den Wert eines vergleichbar großen Webkorpus reduziert. Auch wenn wir uns mit dessen unbekannter Zusammensetzung eine weitere Ungenauigkeits- bzw. Fehlerquelle eingehandelt haben können, haben wir bei unseren Stichproben den Eindruck gewonnen, dass über die Kumulierung der Worthäufigkeiten diese Effekte weitestgehend nivelliert wurden (da nicht unbedingt alle Wortformen eines Paradigmas als regional bewertet wurden).

### 6.3.9 Varianten/Varietäten (regional, diachron, Rechtschreibreform, insbes. Getrennt-/Zusammenschreibung)

Je nachdem, wie weit der Grundformenbegriff ausgelegt werden soll, könnten auch Schreibvarianten unter demselben Lemma subsumiert werden: *daß/dass*, *aufwendig/aufwändig*, usw. Damit öffnet sich ein weites Feld regionaler (Schweizer *dass*) und diachroner Varianten. Auch internet-typische Schreibweisen und die verschiedenen (Akzeptanz-)Phasen der Rechtschreibreform haben einiges zu einer Dynamik bzw. Uneinheitlichkeit des Sprachgebrauchs beigetragen. Gerade auch der Bereich der Getrennt-/Zusammenschreibung ist eng verflochten mit der Frage, welche Einheit als Lemma anzusetzen ist (*sitzenbleiben* vs. *sitzen bleiben* vs. *er blieb sitzen* usw.).

Mit dem letzten Punkt berühren wieder die Fragen, die wir bereits bei den diskontinuierlichen Konstituenten nur ansatzweise aufrollen konnten, sodass dieser Bereich für die aktuelle Version zurückgestellt wurde. In allen anderen Punkten sind wir weitestgehend den Vorschlägen des Lemmatisierers gefolgt, der (allerdings nicht ganz einheitlich) darauf ausgelegt ist, die bundesdeutsche Schreibvariante nach aktueller Rechtschreibregelung zu bevorzugen. In den erkannten Einzelfällen haben wir Abweichungen davon korrigiert. Quantitative Vergleiche zwischen den hier genannten Schreibweisen sind deshalb aber nur schwerlich möglich, da die spezifische Information bei der Kumulierung verlorengegangen sein kann.

### 6.3.10 Kurzwörter

Zu manchen Wortformen gibt es verkürzte Varianten (z.B. *Alu* für *Aluminium*), bei denen sich wie im vorherigen Punkt die Frage stellt, ob sie gemeinsam unter eine Grundform gefasst werden sollen. Gerade darin zeigt sich die Veränderung des Sprachgebrauchs, heute weiß kaum noch ein Sprecher, dass *Bus* nur die Kurzform von *Omnibus* ist. Dieses Phänomen müsste aber sehr wahrscheinlich aus diachroner Perspektive betrachtet werden, interessant wäre hierbei sicher auch, einen Zusammenhang zwischen den Längen der jeweiligen Formen und ihren Gebrauchshäufigkeiten herzustellen.



Für die aktuelle Version wurde dieser Bereich zurückgestellt.

#### 6.3.11 Akronyme, Einzelbuchstaben und Kürzel

Uns erscheint es sinnvoll, Akronyme – ähnlich wie Eigennamen – nicht unbedingt als zum Kernbestand einer Sprache gehörig zu betrachten, stattdessen evtl. in eigenen Listen zu dokumentieren. Bei Einzelbuchstaben/Kürzeln stellt sich die Frage, ob sie eigene Einheiten darstellen oder zu ihrer Vollform expandiert werden sollen (z.B. *z.B.* zu *zum* und *Beispiel?*). Vorläufig möchten wir sie aus unseren Betrachtungen herausnehmen. Deshalb werden mit einer ersten, sehr groben Annäherung Zeichenfolgen, die entweder (a) nur aus einem Zeichen bestehen, die (b) keine Vokale/Umlaute/y enthalten oder (c) einen Majuskel an nicht-erster Stelle enthalten, als Kandidaten für diese Kategorie herausgefiltert. Darüberhinaus wurde diese Fragestellung für die aktuelle Version nicht weiter bearbeitet.

#### 6.3.12 unselbstständige Morpheme

Manche Wörterbücher verzeichnen auch unselbstständige Morpheme (z.B. *-sche*), die aber nur sehr schwierig systematisch zu erfassen sind, sei es, weil sie alle berücksichtigt oder herausgefiltert werden sollen.

Dieser Bereich wurde für die aktuelle Version zurückgestellt.

#### 6.3.13 Verschmelzungen (Amalgamierung)

(*ans, zum, zur, fürs, ?? fortan, infolge, aufgrund, zuhause*)

In Fällen der Verschmelzung stellt sich die Frage, ob die beteiligten Formen (*ans* = *an das*, *zum* = *zu dem*, ...) nicht herausgebrochen werden und entsprechend behandelt werden sollten.

Diese Fragestellung wurde für die aktuelle Version zurückgestellt.

#### 6.3.14 Nennung der Grundform

In den meisten Fällen ist es, wie in der Einleitung bereits kurz erwähnt, plausibel, eine konkrete unmarkierte Form aus dem Paradigma als Grundform anzugeben. Für Nomen ist dies der Nominativ Singular. Einige Spezialfälle, die nur im Singular verwendet werden, stellen keine Problemfälle dar (sog. Singularia Tantum wie z.B. *Hass, Milch, Sand, Getreide, Obst*), im komplementären Fall der Nomen, die nur im Plural vorkommen (sog. Pluralia Tantum wie z.B. *Eltern, Geschwister, Spesen, Ferien*) erscheint es aber intuitiv sinnvoll, von dieser Regelung abzuweichen. Die Frage wird allerdings komplexer, weil analog hinterfragt werden müsste, ob nicht bei allen Paradigmen eine fast ausschließlich (oder sogar generell: die am häufigsten) verwendete Form angegeben werden sollte. Wir haben die in einem Vergleichswörterbuch nicht belegten Formen manuell ausgewertet und in den von uns erkannten Fällen einer fast ausschließlichen Verwendung der Pluralformen ersatzweise deren kumulierte Häufigkeit unter dem Lemma des Nominativ Plurals angegeben.

Eine ähnliche Frage stellt sich auch bei Adjektiven, die gemeinhin im Positiv genannt werden. Diejenigen, die nicht steigerungsfähig sind, stellen wiederum kein Problem dar (*einzig, ewig, erster, letzter (letzterer??)*). Diejenigen allerdings, die nur gesteigert vorkommen (*weltgrößte, dienstältere, dienstälteste, äußerst* (von „*außen*“??)), sollten mittelfristig entsprechend gekennzeichnet werden. Dies haben wir für die aktuelle Version jedoch noch zurückgestellt.

Für manche Wortformenparadigmen halten wir es aber für nahezu kontraintuitiv, eine Form als Nennform auszuwählen, da hierbei die Wahrnehmung des Unterschieds zwischen Wortform und

Grundform unnötig erschwert wird (z.B. *der, die, das* usw.). Wir haben dafür weitestgehend die Ersatzdarstellung des Lemmatisierers übernommen, die die variierenden Bestandteile durch einen Bindestrich ersetzt (z.B. *d-, d-jenig-*). In Einzelfällen wurde diese Ersatzdarstellung von Hand nachgetragen.

### 6.3.15 Movierung

Für Personen-, insbesondere für Berufsbezeichnungen gibt es häufig zwei parallele Formen (männlich/weiblich). Einige Formen lassen sich oberflächlich eindeutig jeweils einem Geschlecht zuordnen (z.B. *Lehrer/Lehrerin, Witwe/Witwer*), einzelne Formen sind aber homonym sozusagen auf beiden Seiten vertreten (*Beamte/Beamter (der/die/eine vs. ein)*). Wir haben den Lemmatisierer in der Einstellung verwendet, die im ersten Fall zwei Lemmata, im zweiten Fall nur ein Lemma angesetzt hat. Mittelfristig wäre eine einheitliche Lösung in der einen oder anderen Form anzugehen.

### 6.3.16 Abgleich mit Wörterbuchlemmastrecken

Grundsätzlich erscheint es natürlich sehr vielversprechend, unsere Grundformenliste mit anderen Lemmalisten abzugleichen. Allerdings gibt es in diesen Listen Lücken und „Leichen“, und der Bezugszeitpunkt kann anders sein: Neueste Formen wurden evtl. nicht berücksichtigt (Neologismen), z.T. liegen andere Prinzipien zugrunde: Fremdwörter, Fachbegriffe, veraltete Schreibweisen, unselbstständige Morpheme, Einzelbuchstaben und Akronyme, verschiedene Epochen gerade während der verschiedenen Phasen der Rechtschreibreform werden je nach Vorgaben unterschiedlich gehandhabt. Leider unterscheiden sich die verschiedenen Listen auch in ihrem Umfang und es liegt kein explizites Relevanzkriterium vor (wie bei uns primär die Häufigkeit), so dass wir die Vergleichslisten nicht sortieren können. Wir wissen also nicht, welche Einträge die ersten 30.000 wären.

	<i>in unserer Liste</i>	<i>nicht in unserer Liste</i>
im Wörterbuch	leichte Bestätigung, könnte im Wörterbuch aber erst viel später berücksichtigt sein	könnte ebenfalls im Wörterbuch weiter hinten zu finden, oder, sofern uns kein Fehler unterlaufen ist, ein veralteter Eintrag (eine „Leiche“) sein
nicht im Wörterbuch	sofern uns kein Fehler unterlaufen ist, könnte dies ein fehlender Eintrag (eine Lücke) sein (*)	---

Die für unsere aktuelle Version relevantesten Erkenntnisse haben wir uns aus der Überprüfung derjenigen Formen erhofft, die zwar in unserer Liste, aber nicht im Wörterbuch zu verzeichnen sind (\*). Diese haben wir manuell in zwei Mengen aufgeteilt: Diejenigen, die aus unserer Sicht einen relevanten Eintrag darstellen, und diejenigen, die als nicht-relevant herauszufiltern sind. Letztere Menge bedarf noch einer detaillierteren Auswertung.

### 6.3.17 Wortreihen

Manche Lemmata (egal welcher Wortklasse) gehören zu kleinen überschaubaren, in sich geschlossenen Reihen: Wochentag- und Monatsbezeichnungen, Tageszeiten, Farben usw. Einerseits ist nicht grundsätzlich anzunehmen, dass alle Elemente einer Reihe ungefähr gleich häufig vorkommen. Es ist z.B. durchaus plausibel, über Primärfarben und im Straßenverkehr und der Politik verwendete Ausprägungen (*rot, gelb, blau, grün*) öfter zu sprechen als über andere; andererseits kann es durch Eigenschaften der Datengrundlage und die Willkür des Einschnitts

(der zumal mitten in einer Häufigkeitsklasse sein) passieren, dass ein Teil der Reihe in unserer Auswahl ist, ein anderer Teil jedoch nicht. Hier ist es überlegenswert, zumindest alle Elemente der Reihe der noch berücksichtigten Häufigkeitsklasse (evtl. sogar einer darüber hinaus) bevorzugt zu berücksichtigen, oder ihre Grenzwertigkeit in einer zusätzlichen Liste festzuhalten. Vorläufig wurde diese Fragestellung aber für die aktuelle Version zurückgestellt.

### 6.3.18 Häufigkeitsklassen

Die Häufigkeit einer Wort- oder Grundform in absoluten Zahlen anzugeben ist wenig sinnvoll. Der Betrachter verbindet damit eine Genauigkeit und eine Zuverlässigkeit der Aussage, die nicht gegeben ist. Aufgrund der Zusammensetzung der Datengrundlage können sich Verzerrungen bei den Wortformfrequenzen ergeben, die oben beschriebenen Problemfelder können zusätzliche Verschiebungen bei den Grundformfrequenzen bewirken. Als relativ stabil und aussagekräftig – gerade auch beim Vergleich unterschiedlich großer Datenbestände – hat sich erwiesen, die Häufigkeit in Form von Häufigkeitsklassen anzugeben. Dabei hat eine Grundform die Häufigkeitsklasse  $N$ , wenn diese Form etwa  $2^N$ -mal seltener vorkommt als die häufigste Form. Für die Grundformenliste ist der Eintrag mit der höchsten Frequenz  $d$ - mit  $f(d)$  = , d.h.

$$N = \text{hk}(\text{lemma}) := \lfloor \log_2(f(d)/f(\text{lemma})) \rfloor$$

also  $f(\text{lemma}) \approx f(d)/2^N$ .

Bsp.

$N =$	0	1	2	3	4	5		10		17
$2^N =$	$2^0$	$2^1$	$2^2$	$2^3$	$2^4$	$2^5$	...	$2^{10}$	...	$2^{17}$
$2^N =$	1	2	4	8	16	32		1024		131072
Bsp.	d-	...	und	von	als	Jahr		aktuell		Lachmuskel

d.h.  $d$ - ist etwa vier Mal häufiger als *und*, etwa acht Mal häufiger als *von* und etwa 131072 Mal häufiger als *Lachmuskel*.

In der veröffentlichten Form ist die Liste auch innerhalb der Häufigkeitsklassen nach der absoluten Häufigkeit sortiert.

### 6.3.19 Qualitätskontrolle

Als Qualitätskontrolle haben wir abschließend verschieden große Mengen an zufällig ausgewählten Einträgen manuell überprüft. Die Ergebnisse dieser Auswertung sind in einen letzten Überarbeitungsschritt eingeflossen.

## 6.4 Zusammenfassung der Entscheidungen für die aktuelle Version

Der aktuellen Version der Wortgrundformenliste liegt eine Wortformenfrequenzliste eines sehr großen Subkorpus des Hauptarchivs des DEUTSCHEN REFERENZKORPUS (DEREKO 2007), das virtuelle Korpus „ccdb-2007“, zugrunde. Daraus wurden alle Wortformen ausgewählt, die nur aus Zeichen des deutschen Alphabets bestehen, maximal 50 Zeichen lang sind und mindestens zweimal vorkommen.

Die Häufigkeiten dieser Wortformen, die (automatisch herausgefiltert und manuell bestätigt) als regional überrepräsentiert erkannt wurden, wurden durch Häufigkeitsangaben aus einem vergleichbar großen Webkorpus ersetzt.

Die Wortformen der überarbeiteten Liste wurden vom Lemmatisierer *g/lemm* (Belica 2004) auf ihre von diesem Verfahren am stärksten favorisierte Grundform zurückgeführt.

In der Menge der vom Lemmatisierer nicht erkannten Wortformen wurden systematisch Kandidaten für neue bzw. unvollständige Adjektivparadigmen sowie im Einzelfall für Verbparadigmen aufgespürt, manuell überprüft und ggf. die Grundform den Wortformen zugeordnet. Bei den übriggebliebenen nicht-lemmatisierten Wortformen wurde die Wortform als Grundform angenommen.

Darauf aufbauend wurde die Häufigkeit einer Grundform über die Kumulation der Häufigkeiten der (automatisch bzw. manuell) zugeordneten Wortformen bestimmt.

Die so berechneten Grundformhäufigkeiten wurden in zwei Bereichen überarbeitet. Bei den Präverbfügungen wurden neue, von einem automatischen Verfahren vorgeschlagene Schätzwerte für die beteiligten Konstituenten manuell überprüft. Außer im Fehlerfall wurden die grundsätzlich bestätigten neuen Werte übernommen; die nur tendenziell akzeptierten Vorschläge wurden in der jeweiligen Richtung mit dem Faktor 2 (Halbierung bzw. „nur“ Verdoppelung) eingearbeitet. Die Menge der Grundformen für Determinantien, bei denen die Angabe der Nennform konfliktträchtig bzw. wenig aussagekräftig sein kann (d-, kein-), wurde ergänzt, bezüglich der kumulierten Wortformen überprüft und im Einzelfall überarbeitet.

Aus der so entstandenen Liste wurden mit einer groben Näherung (Zeichenfolgen aus einem Zeichen bzw. ohne Vokal bzw. mit Großschreibung an nicht-erster Stelle) ansatzweise Akronyme ausgesondert.

Die ersten (nach Grundformhäufigkeit sortierten) 50.000 Einträge aus dieser Liste wurden mit einer Wörterbuchlemmastrecke abgeglichen. Die ersten 7.000 nicht im Wörterbuch belegten Einträge wurden manuell überprüft. Vermeintlich falsche Grundformen aufgrund der falschen Zuordnung zur Wortklasse (Verb anstelle adjektivischem Partizip Präsens; adjektivisches Partizip Perfekt anstelle Verb; Adjektiv bzw. Verb anstelle Substantiv usw.) sowie fehlende Einordnung als Pluralia Tantum wurden manuell überprüft und ggf. überarbeitet. Variierende Schreibweisen (alte/neue Rechtschreibung, Varietäten usw.), sofern nicht bereits vom Lemmatisierer „quasi-normiert“ und sofern nicht im Wörterbuch belegt, wurden ggf. in die bundesdeutsche Schreibweise nach aktuell geltender Rechtschreibung überführt. Eigennamen, geographische (außer vermeintlich „heimatkundlich“ interessante) Bezeichnungen und Fremdwörter, vor allem Anglizismen, wurden zur Aussonderung markiert. Nach deren Filterung verblieben unter den ersten (nach Grundformhäufigkeit sortierten) 30.000 Einträgen nur solche, die entweder im Wörterbuch belegt waren oder manuell bei der Auswertung der im Wörterbuch nicht belegten Einträge als relevant eingestuft wurden.

Bei der abschließenden Qualitätskontrolle wurden vier verschieden große Zufallsstichproben aus dieser Liste (2000, 1000, 500 und 250 Einträge) manuell überprüft. Die wenigen dabei erkannten unerwünschten Einträge wurden zu der Menge der auszusondernden Kandidaten hinzugefügt, unerwartete Partizipien wurden überprüft und je nach Einzelfall überarbeitet. Die bei der Überprüfung erkannten, vom Lemmatisierer (nach heutigem Stand falschen) vorgeschlagenen *ss-* (anstelle *ß-*) Schreibweisen wurden systematisch bei allen Einträgen korrigiert.

Auch nach der Filterung der zusätzlich auszusondernden Kandidaten verblieben in einer neuen

Fassung unter den ersten (nach Grundformhäufigkeit sortierten) 30.000 Einträgen nur solche, die entweder durch den Wörterbuchabgleich oder durch den Bearbeiter legitimiert wurden. Diese 30.000 Grundformen sind mir ihrer Häufigkeitsklasse in der aktuellen DEREWO-Liste angegeben.

## 7 Ausblick

Auf der Grundlage der Erfahrungen, die wir bei dieser Version gesammelt haben, und aufgrund der Rückmeldungen, die uns in der nächsten Zeit erreichen, planen wir eine Neuauflage einer Grundformenliste in einer noch festzulegenden Größenordnung.

## 8 Referenz

Belica, Cyril (1994). A German Lemmatizer. Final Report MLAP93-21/WP2. Luxemburg.

DEREKO (2007): DEUTSCHES REFERENZKORPUS, <http://www.ids-mannheim.de/kl/projekte/korpora/>, Stand: 2007.

## 9 Lizenzbestimmungen

(zu zitieren als:)

*Korpusbasierte Wortgrundformenliste DEREWO, v-30000g-2007-12-31-0.1, mit Benutzerdokumentation,*

<http://www.ids-mannheim.de/kl/derewo/>,

© Institut für Deutsche Sprache, Programmbereich Korpuslinguistik, Mannheim, Deutschland, 2007.

Die Wortgrundformenliste und die Dokumentation bilden eine Einheit. Diese Lizenzbestimmung darf aus keinem der beiden Dokumente entfernt werden.

Dieses Werk ist unter die Creative Commons-Lizenz (by-nc) gestellt (<http://creativecommons.org/licenses/by-nc/3.0/deed.de>).

**Namensnennung – Keine kommerzielle Nutzung 3.0 Unported**

*Sie dürfen:*

- das Werk vervielfältigen, verbreiten und öffentlich zugänglich machen  
Bearbeitungen des Werkes anfertigen

*Zu den folgenden Bedingungen:*

- Namensnennung. Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen (wodurch aber nicht der Eindruck entstehen darf, Sie oder die Nutzung des Werkes durch Sie würden entlohnt).
- **Keine kommerzielle Nutzung.** Dieses Werk darf nicht für kommerzielle Zwecke verwendet werden.
- Im Falle einer Verbreitung müssen Sie anderen die Lizenzbedingungen, unter welche dieses Werk fällt, mitteilen. Am Einfachsten ist es, einen Link auf die o.g. Seite einzubinden.
- Jede der vorgenannten Bedingungen kann aufgehoben werden, sofern Sie die Einwilligung des Rechteinhabers dazu erhalten.

Diese Lizenz lässt die Urheberpersönlichkeitsrechte unberührt.

## 10 Kontakt

Falls Sie speziellere Informationen benötigen, als dieses Werk bereithält, oder Sie dieses Werk über die eingeräumten Rechte hinaus nutzen möchten, wenden Sie sich bitte an [derewo@ids-mannheim.de](mailto:derewo@ids-mannheim.de).

Bei Veröffentlichung auf diesem Werk aufbauender Forschungsergebnisse, bitten wir um eine kollegiale Mitteilung an [derewo@ids-mannheim.de](mailto:derewo@ids-mannheim.de).