

## Lösungen aus Kegli 2 (Korpuslinguistik)

### Kapitel 1

Für die Lösung der Aufgaben 1, 2 und 7 wurden die Suchmaschinen Google und Yahoo verwendet. Gesucht wurde auf allen deutschsprachigen Seiten. Die Ergebnisse stammen vom 7.3.2007.

1. *wegen dem Regen* / *wegen des Regens*:  
Beide Formen werden im Internet verwendet.

	Google	Yahoo
<i>wegen dem Regen</i>	818	2.090
<i>wegen des Regens</i>	17.500	9.720

Die Form *wegen des Regens* ist somit zwischen fünf (Yahoo) und 21 Mal (Google) häufiger als *wegen dem Regen*.

*Praktika* / *Praktikas*: Beide Formen werden im Internet verwendet.

	Google	Yahoo
<i>Praktika</i>	1.950.000	7.070.000
<i>Praktikas</i>	172.000	62.100

Die Pluralform *Praktika* ist in Google etwa elf Mal häufiger belegt als die Form *Praktikas*; in Yahoo findet sich die Form *Praktika* sehr viel häufiger (112 Mal) als *Praktikas*.

2. Befragung:  
Von zehn befragten Personen wählten neun spontan die Partizipform *downgeloadet*, nur eine Person nannte die Form *gedownloadet*.

Internetrecherche:

	Google	Yahoo
<i>downgeloaded</i>	129.000	72.000
<i>Praktikas</i>	172.000	82.100

Während in der gesprochenen Sprache also deutlich die Form *downgeloadet* dominiert, ist im Internet die Form *gedownloadet* etwas häufiger belegt: um ein Drittel häufiger in Google bzw. um ein Sechstel in Yahoo.

3. In einem Korpus zur Fachsprache des Rechts sollten u.a. folgende Textarten enthalten sein:  
Gesetzestexte, juristische Fachzeitschriften, Urteilschriften, Korrespondenz zwischen Kanzleien und Mandanten (Schriftverkehr, Stellungnahmen, juristische Beratungen)

usw.), Protokolle von Gerichtsverhandlungen, juristische Fach- und Lehrbücher, Jura-Vorlesungen an Hochschulen... Je nach Untersuchungsziel können zusätzlich Newsletter von Kanzleien oder allgemeine Presseartikel zu juristischen Themen (Prozessberichterstattung, Gesetzgebung usw.) relevant sein.

Als Verfasser der Texte kommen infrage:

Rechtsanwälte, Richter, Rechtswissenschaftler und andere Juristen, Fachjournalisten; je nach Untersuchung auch andere Verfasser, die sich mit juristischen Themen beschäftigen.

Unterschiede zu einem Korpus der Jugendsprache:

Die Korpora unterscheiden sich hauptsächlich durch die unterschiedlichen Sprechergruppen. Während Jugendsprache primär über das Alter der Sprecher definiert wird und erst sekundär über die Kommunikationssituation, ist die Fachsprache des Rechts stärker an bestimmte Kommunikationssituation (z.B. vor Gericht, im Beruf) gebunden. Das Alter der Sprecher ist hier nicht konstitutiv. Im Gegensatz zur Rechtssprache handelt es sich bei Jugendsprache überwiegend um gesprochene Sprache, weswegen ein Korpus der Jugendsprache unbedingt einen hohen Anteil an gesprochenen Texten enthalten sollte.

Unterscheidung zum Korpus der gesamten deutschen Sprache:

Ein Korpus zur Fachsprache des Rechts (oder der Jugendsprache) bildet eine einzelne Varietät des Deutschen ab. In einem Korpus der gesamten deutschen Sprache hingegen sollten möglichst viele unterschiedliche Varietäten vertreten sein, um die deutsche Sprache möglichst repräsentativ abzubilden. Im Idealfall enthielte ein Korpus der gesamten deutschen Sprache Daten aus allen Varietäten und deren Anteil im Korpus wäre entsprechend zu ihrem Vorkommen in der deutschen Sprache gewichtet.

4. Zur Jugendsprache zählen alle schriftlichen und mündlichen Äußerungen, die innerhalb der Sprechergruppe der Jugendlichen produziert werden, wobei der Begriff "Jugendlicher" konkret definiert werden muss. Beispiele für jugendsprachliche Texte sind Gespräche unter Jugendlichen, SMS, E-Mails, Chats, Schülerzeitungen sowie andere von Jugendlichen gestaltete Texte, die sich an Jugendliche wenden. Kriterien zur Abgrenzung von anderen Texten sind insbesondere das Alter der Sprecher und der bzw. die Kommunikationspartner. Auch die Kommunikationssituation, konkret die An- bzw. Abwesenheit von Erwachsenen, ist zu berücksichtigen.
5. Der Vorteil ist, dass ein solches Korpus elektronisch recherchierbar ist und dennoch schnell (Übernahme aus dem Internet), einfach (kein Einscannen oder Abtippen) und preiswert (keine Kopierkosten) zusammengestellt werden kann. Um die medizinische Fachsprache möglichst breit abzubilden, ist es sinnvoll, das Korpus nicht aus einem einzigen langen Text (z.B. einer Dissertation), sondern aus mehreren kleineren Textproben aufzubauen, die möglichst unterschiedlichen Fachgebieten (z.B. Chirurgie, Innere Medizin, Kinderheilkunde, Augenheilkunde usw.) entstammen sollten. Wird der Aufbau eines Korpus mit minimalem Aufwand angestrebt, sind Quellen wenig attraktiv, die einen hohen Beschaffungs- und/oder Bearbeitungsaufwand erfordern. Dazu gehören an schriftlichen Texten Patientenakten, Arztbriefe und krankenhauserne Schriftstücke, die nicht öffentlich zugänglich sind und für deren Einsichtnahme besondere Genehmigungen erforderlich sind, sowie die meisten gesprochenen Texte, etwa Gespräche von medizinischem Personal oder Patientengespräche. Hier fällt neben dem Einholen von Genehmigungen ein hoher Bearbeitungsaufwand (Aufnahme und Transkription) an.

6. Bei der Auswahl der Texte sollten die Kriterien regionale und soziale Herkunft, Alter, Geschlecht und eventuell die Interessen der jugendlichen Sprecher berücksichtigt werden. Im Korpus sollten alle Textsorten vertreten sein, die Jugendliche produzieren, und zwar in dem Verhältnis, in dem die Texte in der Jugendsprache auftreten. Wichtig ist, dass das Korpus einen hohen Anteil an gesprochener Sprache enthält, da Jugendsprache überwiegend gesprochen wird. Ein repräsentatives Korpus der Jugendsprache würde also z.B. jeweils zur Hälfte Texte von Jungen und Mädchen enthalten. Die Texte würden je zu einem Drittel von jüngeren, mittelalten und älteren Jugendlichen stammen. Die Herkunft der jugendlichen Sprecher würde sich geografisch (z.B. je 20% der Sprecher aus dem Norden, Süden, Westen, Osten und der Mitte des deutschen Sprachgebiets) und sozial gleichmäßig verteilen (z.B. je ein Drittel aus der Unterschicht, der Mittelschicht und der Oberschicht). Neben einem hohen Anteil an gesprochener Sprache (persönliche Gespräche, Telefongespräche, Monologe, Dialoge...) sollten zumindest die wichtigsten jugendsprachlichen Textsorten (E-Mails, SMS, Chats, Briefe, Tagebucheinträge) in das Korpus aufgenommen werden. Die Repräsentativität des Korpus würde noch erhöht, wenn man die einzelnen Kriterien entsprechend den tatsächlichen demografischen Gegebenheiten berücksichtigen würde.
7. Die Autorin kennt und verwendet sowohl Rheinische Verlaufsform als auch doppeltes Perfekt. Verwendet wurden die Suchmaschinen Google und Yahoo. Gesucht wurde auf deutschsprachigen Seiten. Die Ergebnisse stammen vom 7.3.2007. Rheinische Verlaufsform und doppeltes Perfekt: Beide Formen werden im Internet verwendet. Gesucht wurde nach den Formen *ich bin am Schreiben* und *ich habe gelesen gehabt*. Internetrecherche:

	Google	Yahoo
<i>ich bin am Schreiben</i>	192	145
<i>ich habe gelesen gehabt</i>	67	15

Für *ich habe ... gelesen gehabt* finden sich deutlich mehr Treffer, in Google: 908, in Yahoo: 1.670

An Grammatiken geprüft wurden die Duden-Grammatik (2005, 7. Auflage), die *Deutsche Grammatik* von Ulrich Engel (2004), der *Grundriss der deutschen Grammatik* (Band 1) von Peter Eisenberg (2006, 3. Aufl.) sowie die *Deutsche Grammatik* von Gerhard Helbig und Joachim Buscha (18. Aufl., 1998).

Die Rheinische Verlaufsform wird lediglich in zwei Grammatiken kurz erwähnt: unter der Bezeichnung 'Verlaufsform' im Duden (2005, §594) sowie unter der Bezeichnung 'am-Infinitiv' bei Eisenberg (2006:200).

Auch das doppelte Perfekt wird nur in zwei Grammatiken behandelt: zum einen in der Grammatik von Engel (2004) (ca. eine halbe Seite), zum anderen in der Duden-Grammatik. Hier befassen sich insgesamt vier Paragraphen mit Bildung (§658) und Funktion (§§745–746, 769) des doppelten Perfekts.

## Kapitel 2

8. Mögliche Einteilung in Teilkorpora im Korpus "Fachsprache des Rechts" (Aufgabe 3):
- nach Rechtsgebieten (z.B. Arbeitsrecht, Strafrecht, Medienrecht usw.)
  - in gesprochene und geschriebene Fachsprache
  - nach Kommunikationssituation (z.B. vor Gericht, bei der Rechtsberatung)
  - nach juristischen Instanzen (Legislative, Judikative, Exekutive)
  - nach Textsorten (z.B. Gesetze, Urteile, Schriftverkehr, Fachliteratur, Lehrwerke, Ratgeber)
  - nach Adressaten (z.B. Juristen, Laien)

Mögliche Einteilung in Teilkorpora im Korpus "Medizinische Fachsprache" (Aufgabe 5):

- nach Fachrichtungen (z.B. Chirurgie, Orthopädie, Pädiatrie, HNO usw.)
- in gesprochene und geschriebene Fachsprache
- nach Kommunikationssituation (z.B. beim Arzt, im Krankenhaus, auf Fachtagungen)
- nach Textsorten (z.B. Arztbriefe, Fachliteratur, Lehrwerke, Ratgeber, Rezepte)
- nach Adressaten (z.B. Ärzte, Pflegepersonal, Patienten)

Mögliche Einteilung in Teilkorpora im Jugendsprache-Korpus (Aufgabe 6):

- regional
- sozial
- nach Altersstufen
- nach Geschlecht
- nach Bildungsstand
- in gesprochene und geschriebene Jugendsprache
- nach Textsorten (z.B. Briefe, Tagebücher, E-Mails, SMS, Telefongespräche)
- nach Adressaten (z.B. Jugendliche, Erwachsene)

9. Beim ICE handelt es sich um ein einsprachiges vergleichbares Korpus. Obwohl Material aus verschiedenen Ländern erhoben wurde, ist das ICE ein einsprachiges Korpus, da es sich immer um regionale Varietäten einer einzigen Sprache, nämlich dem Englischen, handelt. Das ICE ist ein vergleichbares Korpus, da die Daten aller Teilkorpora zwar nach denselben Kriterien erhoben wurden, die Texte aber unabhängig voneinander entstanden sind. Parallelkorpora hingegen enthalten Originaltexte und deren Übersetzungen. Parallelkorpora sind grundsätzlich mehrsprachig.

### Kapitel 3

Für die Lösung der Aufgaben 13 wurde die Suchmaschine Google verwendet. Gesucht wurde auf allen deutschsprachigen Seiten. Die Ergebnisse stammen vom 3.8.2007.

10. Textwörter: 162

Types (Lemma-Types): 37

Tokens: 47

ARM (1 Token: Arm); AUGUSTINER (1 Token: Augu\_tinern); BANKET (1 Token: Pancket); BEFEHL (1 Token: Befehlig); BEGRÄBNIS (1 Token: begrebnis); BRÜSSEL (1 Token: BrΦ\_\_el); DIENER (3 Tokens: Diener, Diener, Diener); DOLCH (1 Token: Dolchen); Don (1 Token: Don); FLORIS (1 Token: Floris); Freude (1 Token: Freud); GALGEN (2 Tokens: Galgen, Galgen); GÄRTNER (3 Tokens: Gertner, Gertner, Gertner); GEMÄCHTSSTOSS (1 Token: Gemecht\_t=ß); GEWALT (1 Token: gewalt); HAND (1 Token: Hand); HAUPTMANN (3 Tokens: Hauptman, Hauptman, Hauptman); HAUS (1 Token: Hauß); HERR (2 Tokens: Herr, Herrn); HOF (1 Token: Hoff); KLOSTER (1 Token: Clo\_ter); MAULTASCHE (1 Token: Maultaschen); NOTWEHR (1 Token: Nothwehr); OFFIZIER (1 Token: Officirn); RITTER (1 Token: Ritter); RODERICO (1 Token: Roderico); SAAL (2 Token: Saal, Saal); SENTENZ (1 Token: Sententα); SPINNOLA (1 Token: Spinnola); STICH (1 Token: Stich); TAG (1 Token: Tag); TANZ (1 Token: Dantα); TÄTER (2 Tokens: Theter, Theter); VENESI (1 Token: Vene\_i); WISSEN (1 Token: wi\_\_en); ZORN (1 Token: 'orn); ZWÖLFTER (1 Token: 12.)

11. In Korpus B befinden sich die meisten Belege für *Blumentopf*. Bezogen auf eine Millionen Textwörter enthält Korpus B 100 Belege, während Korpus A 80 Belege und Korpus C nur 67 Belege enthält.

Korpus	Belege	Textwörter	Belege je Mio. Textwörter
A	80	1.000.000	80
B	100	1.000.000	100
C	100	1.500.000	67

## 12. Daten normalisiert je Mio. Textwörter

Korpus	-er		-heit/-keit		-um	
	Types	Tokens	Types	Token	Types	Tocken
Dürer-Korpus	211	1.591	173	741	439	5.552
Wissensliteratur	475	4.199	423	6.128	955	4.858

## Type-Token-Verhältnis

Korpus	-er		-heit/-keit		-um	
	Verhältnis		Verhältnis		Verhältnis	
Dürer-Korpus	1 : 7,5		1 : 4,3		1 : 12,7	
Wissensliteratur	1 : 8,8		1 : 14,5		1 : 5,1	

Die meisten Nominalisierungen mit den Suffixen *-er*, *-heit/-keit* und *-ung* finden sich im Korpus der Wissensliteratur, einzige Ausnahme stellen die *-ung*-Tokens dar.

Das Korpus der Wissensliteratur enthält zwei- bis zweieinhalb Mal so viele Types je Suffix wie das Dürer-Korpus. Es enthält etwa zweieinhalb Mal so viele *-er*-Tokens wie das Dürer-Korpus und mehr als achtmal so viele Tokens mit dem Suffix *-heit/-keit*. Die Anzahl der *-ung*-Tokens ist hingegen etwas geringer als im Dürer-Korpus.

Das Type-Token-Verhältnis für das Suffix *-er* ist im Dürer-Korpus und im Korpus der Wissensliteratur annähernd gleich. Bei *-heit/-keit* finden sich im Korpus der Wissensliteratur mehr als dreimal so viele Tokens je Type wie im Dürer-Korpus. Bei den *-ung*-Nominalisierungen hingegen enthält das Dürer-Korpus mehr als doppelt so viele Tokens pro Type wie das Korpus der Wissensliteratur.

## 13. (Angaben nach dem Wörterbuch der deutschen Gegenwartssprache (WDG) 5.4.2007)

Alle drei Wörter haben mehrere Bedeutungen.

*Karte* bezeichnet allgemein ein rechteckiges Stück festes Papier oder Karton, das zu verschiedenen Zwecken verwendet wird. Als *Karte* bezeichnet werden insbesondere: 1. Spielkarte; 2. Post-, Ansichtskarte; 3. Eintrittskarte; 4. Fahrkarte; 5. Visitenkarte; 6. Speisekarte; 7. Lebensmittelkarte; 8. Landkarte.

Das Wort *decken* hat folgende Bedeutungen: 1. etw. auf etw. legen, über etw. breiten; 2. die Nachfrage befriedigen; 3. begatten; 4. etw., jmdn., sich schützen; 5. den gegnerischen Spieler bewachen; 6. (Bankwesen) a) eine Sicherheit für einen Kredit liefern, b) für etw. aufkommen; 7. (*sich decken*) gedanklich oder formal zusammenfallen.

Das Wort *grün* bezeichnet die Farbe des Sonnenspektrums zwischen Gelb und Blau. Daneben hat es eine Reihe von Sonderbedeutungen, die mit der Farbe verbunden sind: a) frisch, saftig, b) unreif, c) Sympathie ausdrückend, d) die Landwirtschaft betreffend.

Häufig verwendet wird *Karte* in Verbindung mit *gut* bzw. *schlecht* (*gute/schlechte Karten haben*), mit *mischen* oder *geben* (*Karten mischen/geben*), mit *schreiben* (*jmdm. eine Karte schreiben*), *bestellen* (*Karten bestellen*) oder *lesen* (*eine Karte lesen*).

An Redewendungen finden sich u.a. *mit offenen Karten spielen*, *seine Karten auf den Tisch legen* oder *alles auf eine Karte setzen*.

Das Verb *decken* steht häufig in Zusammenhang mit *Dach* oder *Tisch* (*ein Dach/den Tisch decken*), mit *Farbe* (*die Farbe deckt gut/schlecht*), mit *Bedarf* bzw. *Nachfrage* (*den Bedarf/die Nachfrage decken*) oder mit *Wechsel* bzw. *Scheck* (*ein (un)gedeckter Wechsel/Scheck*).

An Redewendungen finden sich u.a. *die (kühle) Erde deckt jmdn.* oder *jmdm. den Rücken decken*.

*Grün* tritt häufig auf in Verbindung mit den Wörtern *Gras, Tee, Bohnen, Äpfel* (*grünes Gras/Tee/Bohnen/Äpfel*), mit *Welle* (*grüne Welle*), *Lunge* (*grüne Lunge*) oder *Star* (*grüner Star*).

An Redewendungen lassen sich u.a. nennen: *jmdn. über den grünen Klee loben, grünes Licht geben, grün hinter den Ohren sein* oder *jmdm. nicht grün sein*.

Am häufigsten verwendet wird im Internet (Google) die Wortform (!) *Karte* (ca. 8,1 Mio. Treffer). Die Wortformen *decken* (ca. 2,3 Mio. Treffer) und *grün* (ca. 2,6 Mio. Treffer) sind deutlich seltener.

Laut Frequenzwörterbuch ist *Karte* (40 Tokens) in der gesprochenen Sprache häufiger als *decken* (33 Tokens) und *grün* (17 Tokens) (Ruoff 1981). In der geschriebenen Sprache hingegen ist *decken* (434 Tokens) häufiger als *Karte* (301 Tokens) und *grün* (229 Tokens) (Ortmann 1975).

14. Verwendet wurde das DWDS–Kernkorpus. Untersucht wurden die ersten 50 Treffer des Korpus, die aus den Jahren 1900–1902 stammen. Die Konkordanz zeigt folgende Bedeutungsvarianten:

19 x Teil eines (Gesetzes)Textes  
2 x Unterbrechung (einer Fläche)  
11 x Teil eines Schuhs  
0 x Ablagerung  
18 x Verkauf von Waren und Produkten

15. Verwendet wurde das Mannheimer Morgen–Korpus des IDS. Untersucht wurden jeweils die ersten 50 Treffer des Korpus, die aus dem Jahr 1995 stammen.

Folgende Kollokationen treten auf:

Bei der Wortform *Mann*:

*ihr Mann* (7x), *junger Mann* (3x), *x-jähriger Mann* (4x), *neuer (starker) Mann* (2x), *ihren Mann stehen*(2x), *Mann der Truppe* (2x)

Bei der Wortform *Frau*:

*seine (zweite) Frau* (7x), *Walters Frau* (3x), *junge Frau* (3x), *x-jährige Frau* (4x), *getötete Frau* (2x), *eine Frau geht ihren Weg* (2x), *als erste Frau* (2x)

Das Wort *Mann* tritt insbesondere im Kontext mit Beruf (*Panzergranadier*) und Karriere (*starker Mann in der Sender–Hierarchie, Chef*) bzw. diversen Delikten (als Straftäter, Attentäter, Unfallbeteiligter, Anzeigender) auf. Männer werden als *stark, bedeutend, mit Mumm und Durchsetzungsvermögen* oder *mit Power und Witz* beschrieben. Auf der anderen Seite finden sich auch Männer, die über ihre Frau definiert werden (*Inge Meysel zusammen mit ihrem Mann John Olden*) sowie verlassene, weinende und geistesgestörte Männer in dem Korpusausschnitt.

Das Wort *Frau* tritt ebenfalls im Zusammenhang mit verschiedenen Delikten (als Unfallbeteiligte, Diebin, Mordopfer) auf. Häufig wird *Frau* als weibliche Anrede verwendet (*Frau Paul, Frau Dr. Hovermann*). *Frau* tritt häufig im Kontext mit Familienangehörigen (dem Ehemann, Tochter) auf. Frauen werden häufig als

Beteiligte bei der Handlung eines Mannes genannt (*Schneider und seine Frau, gemeinsam mit seiner Frau*), teilweise wird die Emanzipation von Frauen aber auch ausdrücklich hervorgehoben (*als erste Frau*). Frauen werden eher als Opfer denn als Täter von Straftaten genannt.

16. Die Wortlisten finden Sie unter folgenden Adressen:

<http://wortschatz.uni-leipzig.de/html/wliste.html>

<http://www.ids-mannheim.de/kl/30000wordforms.dat>

Verglichen wurden die ersten 100 Wortformen. Dabei fällt Folgendes auf:

- 97 der ersten 100 Wortformen sind identisch. Folgende drei Wortformen kommen nur beim IDS vor: *dass* (Rang 87), *Er* (Rang 99), *Jahre* (Rang 95). Folgende drei Wortformen kommen nur beim Projekt Deutscher Wortschatz vor: *Millionen* (Rang 97), *sagte* (Rang 100), *zwischen* (Rang 94). Dass die Wortform *dass* beim Projekt Deutscher Wortschatz fehlt, ist darauf zurückzuführen, dass die Wortliste noch vor der Rechtschreibreform (8.1.2001) erstellt wurde.
- Die ersten 10 Wortformen sind – abgesehen von *das* und *mit*, die einmal auf Platz 8, einmal auf Platz 9 stehen, auch in der Rangfolge identisch.
- Insgesamt sind die Abweichungen in der Rangfolge eher gering: Die ersten 30 Wortformen des IDS finden sich unter den ersten 30 Wortformen des Projekts Deutscher Wortschatz. Die einzige Ausnahme ist *Der* (IDS Rang 30), das sich beim Projekt Deutscher Wortschatz auf Rang 35 findet. Insgesamt gibt es nur 11 Wortformen, die in der Rangfolge mehr als 10 Positionen differieren: *diese, dieser, eines, ich, immer, können, Prozent, Uhr, Sie, vor, wieder*. Lediglich bei *Uhr* ist die Differenz größer als 20 Positionen.

17. Gelistet werden orthografische Wortform–Typen

12.	1	Hauptmann	3
aber	1	Hauß	1
abgehawen	1	hernach	2
allen	1	Herr	1
als	1	Herrn	1
alßbald	1	Hoff	1
an	1	ihme	1
Arm	1	im	1
auch	1	in	3
Augustinern	1	ins	1
außgezogen	1	ist	2
Befehlig	1	jederman	1
befohlen	1	kommen	1
begleitet	1	lassen	1
begrebnis	1	man	1
bekleidt	1	Maultaschen	1
Brüssel	1	mit	1

Closter	1	müssen	1
Dantz	1	nachkommen	1
darauß	1	niemand	1
das	1	Nothwehr	1
dem	1	nun	1
Den	1	nur	1
den	6	oder	1
der	6	Officirn	1
des	1	ohn	1
die	2	Pancket	1
Diener	3	rechte	1
diß	1	Ritter	1
Dolchen	1	Roderico	1
Don	1	Saal	2
dritten	1	schlecht	1
ein	5	sein	1
eindringen	1	seinem	2
eingestellt	1	seinen	2
einziehen	1	Sententz	1
entzwischen	1	sich	2
erstochen	1	Spannischer	1
etliche	1	Spinnola	1
Floris	1	Stich	1
frembder	1	Tag	1
Freud	1	Theter	2
für	1	vnd	11
Galgen	2	vnnd	1
geben	1	Venesi	1
gebrauchen	1	von	3
gedünckt	1	weil	1
geführt	1	welchen	1
geführt	1	wie	1
gehalten	1	wissen	1
gehenckt	1	wollen	1
Gemechtstöß	1	worden	2
genagelt	1	Zorn	1
genant	1	zu	2
Gertner	3	zulassen	1
gewalt	1	zum	1
gleich	1	zur	1
Hand	1	zwey	1
hat	1		

## Kapitel 4

### 18. Blaha et al. (2001)

1. Untersuchungsgegenstand: Charakteristika der Verwaltungssprache
2. Varietät: Sprache der Bochumer Stadtverwaltung
3. relevante sprachliche Strukturen und Einheiten: nicht angegeben
4. quantitative oder qualitative Analyse: nicht angegeben (erschließbar: eher qualitativ)
5. Untersuchungszeitraum: nicht angegeben (erschließbar: Gegenwartssprache)
6. synchrone oder diachrone Analyse: nicht angegeben (erschließbar: synchron)

### O'Halloran (2002)

1. Untersuchungsgegenstand: Verbreitung von Fremdwörtern im Deutschen
2. Varietät: Modesprache (Frauenzeitschrift *Brigitte*) und Standardsprache (*Stern*, *Berliner Illustrierte Zeitung*)
3. relevante sprachliche Strukturen und Einheiten: Gallizismen und Anglizismen
4. quantitative oder qualitative Analyse: quantitativ
5. Untersuchungszeitraum: 1902–1997
6. synchrone oder diachrone Analyse: diachron

### Steyer (2002)

1. Untersuchungsgegenstand: Rekonstruktion von sprachlichem Wissen über den Begriff *Hund*, dessen Bedeutung und Verwendung im Deutschen
2. Varietät: Standardsprache (IDS-Korpora)
3. relevante sprachliche Strukturen und Einheiten: Kontexte, die das Wort *Hund* enthalten
4. quantitative oder qualitative Analyse: qualitativ (Kollokationsanalyse)
5. Untersuchungszeitraum: überwiegend Gegenwartssprache (IDS-Korpora)
6. synchrone oder diachrone Analyse: nicht angegeben (erschließbar: synchron)

### 19.

- a. Die\_DIE Katze\_KATZE hat\_HABEN viele\_VIEL Mäuse\_MAUS gefangen\_FANGEN
- b. Er\_ER war\_SEIN mit\_MIT seinen\_SEIN Freunden\_FREUND in\_IN der\_DER Oper\_OPER
- c. Bist\_SEIN du\_DU gestern\_GESTERN nach\_NACH dem\_DEM Kino\_KINO bei\_BEI Claudia\_CLAUDIA gewesen\_SEIN

### 20. Tags für Tempus: Präsens (PRÄ), Präteritum (PRT)

- a. Die\_ART-DEF\_NOM\_SIN\_FEM  
Katze\_SUB\_NOM\_SIN\_FEM  
hat\_VER-AUX\_3\_SIN\_PRÄ  
viele\_ADJ\_AKK\_PLU\_FEM  
Mäuse\_SUB\_AKK\_PLU\_FEM  
gefangen\_PA2

- b. Er\_PER\_3\_SIN\_NOM\_MAS  
 war\_VER\_3\_SIN\_PRT  
 mit\_PRP  
 seinen\_POS\_DAT\_PLU\_MAS  
 Freunden\_SUB\_DAT\_PLU\_MAS  
 in\_PRP  
 der\_ART-DEF\_DAT\_SIN\_FEM  
 Oper\_SUB\_DAT\_SIN\_FEM
- c. Bist\_VER-AUX\_2\_SIN\_PRA  
 du\_PER\_2\_SIN\_NOM  
 gestern\_ADV  
 nach\_PRP  
 dem\_ART-DEF\_DAT\_SIN\_NEU  
 Kino\_SUB\_DAT\_SIN\_NEU  
 bei\_PRP  
 Claudia\_SUB\_DAT\_SIN\_FEM  
 gewesen\_PA2

21.

- a. Token: *Anhengern*, Type: ANHÄNGER  
 b. Token: *Spiller*, Type: SPIELER  
 c. Token: *Unternehmer, Orgelspieler*; Type: UNTERNEHMER, ORGELSPIELER  
 d. Token: *Anhänger*; Type: ANHÄNGER

Probleme bei Zuordnung:

- Schreibvarianten (*Anhengern, Anhänger*)
- bei Anhänger handelt es sich in Satz a um eine Person (*Anhänger=Gefährte*), in Satz d um ein Objekt (*Anhänger=Wagen*)
- es ist nicht klar, ob *Spieler* in Satz b und *Orgelspieler* in Satz c einem gemeinsamen Lemma SPIELER oder zwei einzelnen Lemmas (SPIELER, ORGELSPIELER) zugeordnet werden müssen

22.

a.

Lemma	<b>ANHÄNGER 1</b>
Basis	anhäng
Morpheme	an+häng+er
Bedeutung	Person
Basis Wortart	Verb
Basis Komplexität	komplex
Beleg	<b>Anhengern</b>
Quellenangabe	Schnitt 1609, Ausgabe 30, Zusatz –, Seite 7, Zeile 11, Spalte –
Kontext	Sambstags hat der Babst etlich schreiben / in Teutschland / an Keys. M. Jhr Durchl. zu Greetz /

	Bayern und andere catholische Fürsten abgehen lassen / sich den Protestirenden und andern <b>Anhengern</b> zu widersetzen

b.

Lemma	<b>SPIELER</b>
Basis	spiel
Morpheme	spiel+er
Bedeutung	Person
Basis Wortart	Verb
Basis Komplexität	einfach
Beleg	<b>Spiller</b>
Quellenangabe	Schnitt 1700, Ausgabe 42, Zusatz Extrablatt, Seite 5, Zeile 19
Kontext	wie sie dann verschidene Personen wegen dero Spillen und Fluchen ergriffen / da dann die <b>Spiller</b> nach dem Zuchthause / die andere aber mit Geld gestraft

c.

Lemma	<b>UNTERNEHMER</b>
Basis	unternehm
Morpheme	unter+nehm+er
Bedeutung	Person
Basis Wortart	Verb
Basis Komplexität	komplex
Beleg	<b>Unternehmer</b>
Quellenangabe	Schnitt 1850, Ausgabe 220, Zusatz 2. Beilage, Seite 2, Zeile 65, Spalte 1
Kontext	Der <b>Unternehmer</b> zeigte sich bei dieser Gelegenheit abermals nicht nur als einen der geübtesten Orgelspieler, sondern auch als einen vorzüglichen Componisten.

Lemma	<b>ORGELSPIELER</b>
Basis	Orgel spiel
Morpheme	orgel+spiel+er
Bedeutung	Person
Basis Wortart	verbale Wortgruppe
Basis Komplexität	komplex
Beleg	<b>Orgelspieler</b>
Quellenangabe	Schnitt 1850, Ausgabe 220, Zusatz 2. Beilage, Seite 2, Zeile 65, Spalte 1
Kontext	Der Unternehmer zeigte sich bei dieser Gelegenheit abermals nicht nur als einen der geübtesten <b>Orgelspieler</b> , sondern auch als einen vorzüglichen Componisten.

d.

Lemma	<b>ANHÄNGER 2</b>
Basis	anhäng
Morpheme	an+häng+er
Bedeutung	Objekt
Basis Wortart	Verb
Basis Komplexität	komplex
Beleg	<b>Anhänger</b>
Quellenangabe	Schnitt 1950, Ausgabe 18, Zusatz –, Seite 11, Zeile 116, Spalte 3
Kontext	Vorläufig kommen vier moderne Omnibusse zum Einsatz, die bei Bedarf mit <b>Anhänger</b> versehen werden

23.

	1609	1650	1700	1750	1800	1850	1900	1950	2000
Produktivität	0,08	0,06	0,08	0,08	0,12	0,12	0,14	0,15	0,22
Type– Token– Verhältnis	4,35	3,68	3,63	3,25	2,71	2,69	2,92	2,72	2,47

Die Produktivität ist 1609 bis 1750 konstant (Ausnahme: Rückgang 1650). Sie steigt 1800 an und bleibt 1850 konstant. Ab 1900 steigt die Produktivität kontinuierlich. Im Jahr 2000 erreicht die Produktivität ihren Höchststand.

Die Zahl der Tokens je Type nimmt zwischen 1609 und 1850 kontinuierlich ab. Sie steigt 1900 nochmals an, fällt dann jedoch wieder ab und erreicht 2000 ihren Tiefststand.

Das Ansteigen der Produktivität und das fallende Type–Token–Verhältnis stehen im Einklang mit der Beobachtung, dass sich das Wortbildungsmuster diachron ausbreitet.

24.

	Erlanger Korpus	Würzburger Korpus	Mainzer Zeitungskorpus	
			Maximum	Minimum
Types	21	48	558	97
Tokens	159	420	1.376	357
Type– Token– Verhältnis	1 : 7,5	1 : 8,8	1 : 2,5	1 : 3,7

Die Zahl der Types und Tokens ist in den beiden Korpora zum Frühneuhochdeutschen jeweils geringer als in jedem einzelnen Teilkorpus des Mainzer Zeitungskorpus.

Das Type–Token–Verhältnis liegt jedoch in beiden Korpora noch über dem Maximum des Mainzer Zeitungskorpusses. Die zwischen dem Frühneuhochdeutschen (Erlanger Korpus, Würzburger Korpus) und dem Neuhochdeutschen (Mainzer Zeitungskorpus) fallende Zahl von Tokens je Type bestätigt die diachrone Ausbreitung des Wortbildungsmusters.

25. Nach der Normalisierung auf 100.000 Textwörter ergeben sich folgende Zahlen:

	1609	1650	1700	1750	1800	1850	1900	1950	2000
Person Types	155	89	159	201	212	288	298	365	<b>500</b>
Person Tokens	647	308	586	669	601	780	948	1.051	<b>1.266</b>
Objekt Types	5	8	14	18	19	18	49	<b>50</b>	41
Objekt Tokens	53	49	48	53	24	51	78	<b>83</b>	58

Die meisten Personenbezeichnungen finden sich mit 500 Types und 1.266 Tokens je 100.000 Textwörter im Teilkorpus 2000. Die meisten Objektbezeichnungen finden sich mit 50 Types und 83 Tokens je 100.000 Textwörter hingegen im Teilkorpus 1950.

26.

Type–Token–Verhältnis	1609	1650	1700	1750	1800	1850	1900	1950	2000
Person	4,18	3,48	3,68	3,32	2,84	2,71	3,18	2,88	2,53
Objekt	10,40	6,00	3,36	2,84	1,26	2,75	1,57	1,66	1,43
ges. (Aufg. 23)	4,35	3,68	3,63	3,25	2,71	2,69	2,92	2,72	2,47

Die Type–Token–Verhältnisse für Personen– und Objektbezeichnungen weichen von dem Type–Token–Verhältnis des Wortbildungsmusters in Aufgabe 23 ab. Bis zum Jahr 1650 liegt die Zahl der Tokens je Type bei den Objektbezeichnungen über, bei den Personenbezeichnungen unter dem Type–Token–Verhältnis des Wortbildungsmusters. Ab dem Jahr liegt die Zahl der Tokens bei den Objekten unter, bei den Personen über dem Type–Token–Verhältnis des Wortbildungsmusters. Jedoch zeigen alle drei Type–Token–Verhältnisse dieselbe – nämlich eine fallende – Tendenz. Daraus lässt sich zweierlei folgern: Erstens steigt diachron die Produktivität der *-er*-Derivation, der *-er*-Personenbezeichnungen und der *-er*-Objektbezeichnungen. Zweitens ist die Bildung von Objektbezeichnungen ab 1700 produktiver, die von Personenbezeichnungen weniger produktiv als das Wortbildungsmuster insgesamt.

## Kapitel 5

Bei den Aufgaben 28–31 zum DWDS–Kernkorpus werden nur die "anzeigbaren Treffer" genannt, da bei der Gesamttrefferzahl der Jahresfilter nicht funktioniert! Die Ergebnisse für die Aufgabe 27 stammen vom 27.9.2007, die Ergebnisse für die Aufgaben 28–34 vom 28.9.2007.

27. Als Sprache angegeben wurde deutsch. Nicht berücksichtigt wurden Wörterbucheinträge, Bestandteile von Derivaten und Komposita sowie URLs. Die Ergebnisse stammen vom 27.9.2007.

Die Konkordanz für *Mittel* zeigte folgende Bedeutungsvarianten:

- Mittelmaß, im Sinne von arithmetisches Mittel (16/50)
- Geld, im Sinne von Finanzmitteln oder flüssigen Mitteln (4/50)
- Methode, (Straf)Maßnahme (3/50)
- rhetorisches Mittel (4/50)
- Werkzeug, im Sinne von Hilfsmittel (8/50)
- Werkzeug, im Sinne von Gegenmittel (3/50)
- Waffen, im Sinne von militärische Mittel (3/50)
- Medikament, Wirkstoff, im Sinne von Arzneimittel (9/50)

Die Konkordanz für *Schein* zeigte folgende Bedeutungsvarianten:

- Illusion, im Sinne von Anschein (4/50)
- Fahrerlaubnis, im Sinne von Führerschein (11/50)
- Name (einer Person/ Musikband) (28/50)
- Banknote, im Sinne von Geldschein (2/50)
- Wertpapier, im Sinne von Optionsschein (4/50)
- Ruhm, im Sinne von Glorienschein (1/50)

28. *Kontos*: 53 Treffer, jedoch nur Genitiv Singular

*Konti*: 12 Treffer, alle Plural. – Zeit: Die Formen stammen überwiegend aus den 30–er–Jahren (6/12).

Die restlichen 6 Formen verteilen sich auf den Zeitraum zwischen 1919 und 1977. – Textsorte: Alle Formen entstammen der Textsorte Zeitung.

*Konten*: 348 Treffer, alle Plural. Die Form findet sich in allen Zeiten und Textsorten. –

Zeit: Die meisten Formen finden sich in den 20–er– und 40–er–Jahren (88 bzw.

54/348). Zwischen 1900 und 1910 sowie zwischen 1951 und 1980 sind die Formen

eher selten (11–22 Formen/Jahrzehnt). – Textsorte: Die Formen finden sich

überwiegend in den Textsorten Zeitung (143/348) und Gebrauchsliteratur (128/348).

In Wissenschaft (55/348) und Belletristik (22/348) sind sie deutlich seltener.

29.

Suchausdruck	
Haus	36.628 Treffer, die sich auf die Wortformen <i>Haus</i> , <i>Hause</i> , <i>Hauses</i> , <i>Häuser</i> und <i>Häusern</i> verteilen
@Haus	14.432 Treffer für die Wortform <i>Haus</i>
bio*	2.773 Treffer, insbesondere Wortformen von <i>biologisch</i> und <i>biographisch</i> , auch Komposita wie <i>biologisch-anthropologisch</i>
Bio*	3.790 Treffer, z.B. <i>Biograph</i> , <i>Biophysiker</i>
*lich	494.783 Treffer, meist adjektivische <i>-lich</i> -Derivate ( <i>nützlich</i> , <i>förderlich</i> , <i>südlich</i> ), zum Teil aber auch Verbformen ( <i>glich</i> , <i>verblich</i> )
"das weiße Haus"	11 Treffer für die Wortfolgen <i>das weiße Haus</i> , <i>des weißen Hauses</i> , <i>dem weißen Hause</i> , <i>die weißen Häuser</i> (die Suche nach "das Weiße Haus" liefert 248 Treffer!)
Mutter&&Vater	2.611 Treffer, Aufzählung oder Gegenüberstellung von Wortformen der Lexeme <i>MUTTER</i> und <i>VATER</i> im selben Satz, z.B. <i>von europäischen Vätern und japanischen Müttern</i> oder <i>Unterschrift des Vaters</i> oder <i>der Mutter</i>
Mutter    Vater	41.777 Treffer, d.h. Sätze, in denen eine flektierte Form von <i>MUTTER</i> ( <i>Mutter</i> , <i>Muttern</i> , <i>Mutters</i> , <i>Müttern</i> ) oder von <i>VATER</i> ( <i>Vater</i> , <i>Vaters</i> , <i>Väter</i> ) vorkommt
Mutter&&!Vater	19.998 Treffer, d.h. Sätze in denen eine flektierte Form von <i>MUTTER</i> ( <i>Mutter</i> , <i>Muttern</i> , <i>Mutters</i> , <i>Müttern</i> ) vorkommt, aber nicht von <i>VATER</i>
“weiß#1Haus“	47 Treffer, bei denen maximal ein weiteres Wort zwischen der Wortform von <i>WEIß</i> und <i>HAUS</i> steht; meist folgen diese direkt aufeinander ( <i>weiße Häuser</i> , <i>dem weißen Haus</i> ), selten steht ein Wort dazwischen ( <i>dem weißen kleinen Haus</i> )

30.

Suchausdruck	
a. Buch&&lesen	483 Treffer (Gesamtkorpus: 1.307)
b. " groß #1 Auge "	477 Treffer, die eine Wortform von <i>GROß</i> und <i>AUGE</i> enthalten und zwischen denen maximal ein weiteres Wort steht, z.B. <i>mit großen Augen</i> , <i>große tiefblaue Augen</i> (Fälle wie <i>das große braune Auge</i> müssen manuell ausgeschlossen werden!)
c.1. gehen	Lexem <i>GEHEN</i> : 1.702 Treffer
c.2. @gehen	Wortform <i>gehen</i> : 370 Treffer
c.3. @ging	Wortform <i>ging</i> : 314 Treffer
c.4. @gegangen	Wortform <i>gegangen</i> : 53 Treffer
d. Bundesrepublik &&!Deutschland	3.970 Treffer (Gesamtkorpus: 5.390)
e.1. dy*	157 Treffer (Gesamtkorpus: 1.688), z.B. zu den Lexemen <i>DYNAMISCH</i> , <i>DYNASTISCH</i>
e.2. *dy	222 Treffer (Gesamtkorpus: 3.813), z.B. <i>Lady</i> , <i>Dandy</i> , insbesondere aber Namen ( <i>Kennedy</i> , <i>Hardy</i> )

31. Die zehn häufigsten Kollokationen sind:

zu *Hund*: Katz, bellen, Hund, klein, Pferd, arm, tot, groß, jung, kommen

zu *Katze* (Suchausdruck Katz verwenden!): Hund, Maus, Katz, Sack, schwarz, jung, springen, Brei, Vogel, Dr.

zu *Maus*: Ratte, Katz, Tastatur, weiß, Kathrein, 2, 7, Maus, klein, 4

Auffällig ist:

- Alle drei Suchausdrücke finden sich Kollokationspartner ihrer selbst (*Hund* mit *Hund*, *Katz* mit *Katz*, *Maus* mit *Maus*) und von einander (*Hund* mit *Katz*, *Katze* mit *Hund* und *Maus*, *Maus* mit *Katz*).
- Viele Kollokationen deuten auf Idiome und Sprichwörter hin: *wie Hund und Katz sein*, *ein armer Hund sein*, *auf den Hund kommen*, *Katz und Maus spielen*, *die Katze im Sack kaufen*, *die schwarze Katze*, *wie die Katze um den heißen Brei schleichen*
- Einige Kollokationen zeigen stereotype Merkmale von Hund (bellen, groß, jung), Katze (schwarz, springen, Maus/Vogel fangen) und Maus (weiß, klein) an, andere geben Tiere oder Gegenstände an, mit denen sie häufig vorkommen (Hund/Pferd, Maus/Ratte, Maus/Tastatur).
- Einige Kollokate – insbesondere bei Maus – scheinen Korpusartefakte zu sein: *Katz/Dr.*, *Maus/Kathrein*, *Maus/2*, *Maus/4*, *Maus/7*.

32. Bei der Suche nach dem Wortende *\*chen* wurden insgesamt 799 Wortformen gefunden.

Die ersten 20 lauten:

*Aachen*, *abbrachen*, *abbrechen*, *abenteuerlichen*, *abgebrochen*, *abgestochen*, *abgestrichen*, *abgewaschen*, *abgestochen*, *abkriechen*, *abwaschen*, *abweichen*, *abwischen*, *abzubrechen*, *abzustecken*, *abzuweichen*, *abzuwischen*, *ähnlichen*, *altmodischen*, *Alvenslebischen*.

Es fällt auf, dass sich unter den 20 Wortformen kein einziges Derivat mit dem Suffix *-chen* findet. Lediglich zwei sind Substantive, wovon eines ein Eigenname (*Aachen*) und eines ein flektiertes substantiviertes Adjektiv (*Alvenslebischen*) ist. Bei drei Wortformen handelt es sich um flektierte Adjektive (*abenteuerlichen*, *ähnlichen*, *altmodischen*), die restlichen 15 sind Verbformen, z.B. Infinitive (*abbrechen*, *abzubrechen*), Präteritalformen (*abbrachen*) oder Partizipien (*abgebrochen*).

33. Die Tokenfrequenz wird maximal für die ersten fünf Types angegeben. Für die Suche nach Affixen muss unter den OPTIONEN für die Lemmatisierung sowohl "Komposita" als auch "Sonstige Wortbildungsformen" aktiviert werden.

Suchausdruck	Types	(Tokens)
Haus	2	Haus (789), haus (1)
&Haus	5	Haus (789), Hause (819), Häuser (106), Häusern (65), Hauses (149)
be*	4.140	be (6), Be (1), beabsichtige (2), beabsichtigen (1), beabsichtigst (1)
*chen	3.715	Aachen (11), abbrachen (2), abbrechen (8), Abendstündchen (1), Abendhandtäschchen (1)
se?en	4	Segen (39), sehen (1.504), Sehen (16), seien (323)
se+en	5	Seen (19), Segen (39), sehen (1.504), Sehen (16), seien (323)
se*en	258	Se-he-nen (1), sechsen (1), Sechsjährigen (2), Sechsschüssigen (2), sechssitzigen (1)
&be-	2.476	abbekamen (1), abbekommen (1), abbestellt (1), Abbestellung (1), Abfallbehälter (2)
&-chen	310	Abendkleidchen (4), Abendplätzchen (1), Alterchen (1), Altlegitimitätsrestchen (1), Altweiberchen (1)
\?	1	? (10.243)

34.

Suchausdruck	Types	(Tokens)
a. Praktikas	1	(40 )
b.1. &downloaden	6	downgeloadet (15), download (65), Downloaden (86), downloaden (115), downloadet (2), downzuloaden (8)
b.2. &updaten	10	UPDATE (25), UpDate (42), Update (963), update (119), Updaten (13), updaten (27), updatet (1), Upgedatet (1), upgedatet (15), upzudaten (5)
c. ein*ung	131	z.B. Ein–Mann–Abfertigung (1), Ein–Mann–Lenkung (1), Einarbeitung (21), Einäscherung (9), Einbehaltung (1)
d. *linguistik*	103	z.B. Areallinguistik (2), Balkan–Linguistik (1), Balkanlinguistik (1), Betriebslinguistik (1), Biolinguistik (1)

zu b: Flexionsparadigma (belegte Formen)

1. Verb *downloaden*

Infinite Formen:

Infinitiv	zu–Infinitiv	Partizip Perfekt
downloaden	downzuloaden	downgeloadet

Finite Verbformen:

	Präsens Indikativ Aktiv	Perfekt Indikativ Aktiv	Futur Indikativ Aktiv	Präsens Indikativ Passiv
1. Sg.		ich habe downgeloadet		
2. Sg.				
3. Sg.	man downloadet	man hat downgeloadet	sie wird downloaden	
1. Pl.				
2. Pl.				
3. Pl.	sie downloaden			sie werden downgeloadet

Daneben wird der Infinitiv *downloaden* in Modalverbkonstruktionen mit *können*, *lassen*, *müssen* und *wollen* verwendet.

2. Verb *updaten* (belegte Formen)

Infinite Formen:

Infinitiv	zu-Infinitiv	Partizip Perfekt
updaten	upzudaten	upgedatet

Finite Verbformen:

	Präsens Indikativ Aktiv	Perfekt Indikativ Aktiv	Futur Indikativ Aktiv	Präsens Indikativ Passiv	Präteritum Indikativ Passiv
1. Sg.				es wird upgedatet	
2. Sg.					
3. Sg.	man updatet	es hat upgedatet			es wurde upgedatet
1. Pl.			wir werden updaten		
2. Pl.					
3. Pl.	sie updaten			sie werden upgedatet	

Daneben wird der Infinitiv *updaten* in Modalverbkonstruktionen mit *können*, *lassen*, *müssen* und *wollen* verwendet.