

# Der lexikographische Prozess im deutschen Wiktionary



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

**Christian M. Meyer**

DFG-Forschungsnetzwerk „Internetlexikografie“  
4. Arbeitstreffen, Trier, 22.–23. November 2012.

Wikiwörterbuch  
**Wiktionary**  
[ˈvɪkʃəˌnɛʀi], *n*  
Das freie Wörterbuch  
ein Wiki-basiertes  
freies Wörterbuch

Hauptseite  
Themenportale  
Zufällige Seite  
Inhaltsverzeichnis

▼ Mitarbeit  
Eintrag erstellen  
Autorenportal  
Wunschliste  
Literaturliste  
Letzte Änderungen

► Hilfe  
► Werkzeuge

▼ In anderen Sprachen  
Afrikaans  
bosanski  
català  
česky  
Cymraeg  
dansk  
Ελληνικά  
English  
Esperanto  
español

Eintrag [Diskussion](#) [Lesen](#) [Bearbeiten](#) [Versionsgeschichte](#)

## Wörterbuch

Ihre [Spenden](#) helfen, Wiktionary zu betreiben.

Wörterbuch ([Deutsch](#)) [\[Bearbeiten\]](#)

**Substantiv, n** [\[Bearbeiten\]](#)

### Worttrennung:

Wör-ter-buch, Plural: Wör-ter-bü-cher

### Aussprache:

IPA: [ˈvœʁtəbuːχ], Plural: [ˈvœʁtəˌbyːçə]

Hörbeispiele:  [Wörterbuch \(Österr.\)](#) <sup>(Info)</sup>, Plural:  [Wörterbücher \(Österr.\)](#) <sup>(Info)</sup>

### Bedeutungen:

[1] Nachschlagewerk für die Schreibweise, Bedeutung, Grammatik, Geschichte und/oder Übersetzung von Wörtern

### Herkunft:

Verdeutschung von *dictionarium* bzw. *lexicon* im 17. Jahrhundert durch [Comenius](#)<sup>[1]</sup>, ursprünglich holländisch: *wordboek*.

### Synonyme:

[1] *manchmal auch:* [Diktionär](#), [Enzyklopädie](#), [Lexikon](#), [Thesaurus](#)

### Gegenwörter:

[1] [Bibliographie](#), [Einführung](#)

### Oberbegriffe:

[1] [Nachschlagewerk](#), [Buch](#)

### Unterbegriffe:

[1] [Abkürzungswörterbuch](#), [Bedeutungswörterbuch](#), [Bildwörterbuch](#), [Dialektwörterbuch](#), [etymologisches Wörterbuch](#), [Fachwörterbuch](#), [Fremdwörterbuch](#), [Häufigkeitswörterbuch](#), [Idiotikon](#), [Mundartwörterbuch](#), [Online-Wörterbuch](#), [Printwörterbuch](#), [Rechtschreibwörterbuch](#), [Rechtswörterbuch](#), [Reimwörterbuch](#), [rückläufiges Wörterbuch](#), [Sachwörterbuch](#), [Schimpfwörterbuch](#), [Schlagwörterbuch](#), [Stilwörterbuch](#), [Synonymwörterbuch](#), [Valenzwörterbuch](#)

### Beispiele:

[1] Er nutzt ein *Wörterbuch* für Korrekturen.



Kasus	Singular	Plural
<b>Nominativ</b>	das Wörterbuch	die Wörterbücher
<b>Genitiv</b>	des Wörterbuchs des Wörterbuches	der Wörterbücher
<b>Dativ</b>	dem Wörterbuch dem Wörterbuche	den Wörterbüchern
<b>Akkusativ</b>	das Wörterbuch	die Wörterbücher

# Der lexikographische Prozess



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

**1. Vorbereitungsphase**

**2. Phase der Datenbeschaffung**

**3. Phase der Datenaufbereitung**

**4. Phase der Datenauswertung**

**5. Phase der Satz- und Druckvorbereitung**

Wiegand et al. (2010)

# 1. Vorbereitungsphase



- „Das Wikiwörterbuch ist das **deutschsprachige** Wiktionary: ein **frei verfügbares**, mehrsprachiges Wörterbuch für den **Wortschatz aller Sprachen**.“
- „[...] nicht von einer festen, bezahlten Redaktion geschaffen, sondern ist vielmehr das **Werk freiwilliger Autoren**.“
- „[...] kein Forum zur Streuung von Wortschöpfungen oder von Wörtern, die in der Standard- oder Umgangssprache sehr selten benutzt werden. Dazu gehören Wörter, zu denen es keine **nachprüfbaren, verständlichen Belege** gibt. Eine Person, die das Wikiwörterbuch benutzt, sollte sich auch **darauf verlassen** können, dass die hier gefundenen Wörter wirklich **geläufig** sind.“

<https://de.wiktionary.org/w/index.php?title=Wiktionary:Hauptseite&oldid=2431424> (Version vom 15.06.2012)

[https://de.wiktionary.org/w/index.php?title=Wiktionary:%C3%9Cber\\_das\\_Wiktionary&oldid=2708038](https://de.wiktionary.org/w/index.php?title=Wiktionary:%C3%9Cber_das_Wiktionary&oldid=2708038) (Version vom 21.10.2012)

[https://de.wiktionary.org/w/index.php?title=Wiktionary:Was\\_Wiktionary\\_nicht\\_ist&oldid=2175236](https://de.wiktionary.org/w/index.php?title=Wiktionary:Was_Wiktionary_nicht_ist&oldid=2175236) (Version vom 12.04.2012)

## Versuch einer systematischen Einordnung nach...

- **Wb-Gegenstandsbereich:** sämtliche Einzelsprachen
- **Wb-Gegenstand:** allgemeines, mehrsprachiges Wb
  - polyinformativ und polysektiv
  - (hauptsächlich) gegenwartsbezogene Sprache
  - „Standardsprache“, z.Tl. Dialekte, Fachvokabular
  - zu den Schwerpunkten zählen Übersetzungen
- **Wb-Form:** polyakzessives Wb
  - alphabetisch (über Index)
  - Sachgruppen (über Themenseiten)
  - onomasiologisch (über Thesaurus-Seiten)

## Versuch einer systematischen Einordnung nach...

- **Publikationsmedium:** Digitales Wb/Online-Wb
- **Lexikographischem Prozess:** gemeinschaftlich bearbeitetes AusbauWb
- **Semiotischer Kodierung:** multimediales Wb
- **Benutzerbezug:** ?
  - Adressatenkreis: nicht angegeben
  - Wörterbuchfunktion: nicht angegeben
  - Sowohl Rezeptions-, Herübersetzungs-, Produktions- und Hinübersetzungsfunktion strukturell möglich

# Lexikographische Instruktionen



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- Einführung für neue Autoren
- 16 Artikelvorlagen nach Wortart
- Hilfeseiten (allgemein und gezielt zu Vorlagen)
- Beispielartikel

## Kollaboratives Prinzip:

„Diese Formatvorlage **muss nicht** beim Neuanlegen bzw. ersten Editieren **komplett ausgefüllt** werden. [...] Wer sich unsicher ist oder irgendetwas nicht weiß, lässt es einfach offen. Nachfolgende Bearbeiter können dies später auch noch eintragen.“

[http://de.wiktionary.org/w/index.php?title=Hilfe:Verwenden\\_der\\_Formatvorlage&oldid=2433266](http://de.wiktionary.org/w/index.php?title=Hilfe:Verwenden_der_Formatvorlage&oldid=2433266) (Version vom 17.06.2012)

# Mikrostrukturprogramm 2004



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

{Sprache}

{Wortverwendung}

**KFZ (Kraftfahrzeug), *n***

**Aussprache:**

**Herkunft:**

**Bedeutung:** Fahrzeug mit eigenem Antrieb

{Bedeutung des Wortes}

**Synonyme:** [Automobil](#) {Bedeutungsgleiche oder bedeutungsverwandte Wörter}

**Gegenteile:**

**Oberbegriffe:** [Fahrzeug](#), [Fortbewegungsmittel](#)

**Unterbegriffe:** [Motorrad](#), [Personenkraftwagen \(PKW\)](#), [Lastkraftwagen \(LKW\)](#), [Bus](#)

**Beispiele:**

**Redewendungen:**

**? Referenzen und weiterführende Informationen:**

**Abgeleitete Begriffe:**

**Übersetzungen:**

- **Englisch:** [Automobile \(English\)](#) Die Interlanguage-Links sollten den Sprachnamen des jeweiligen Wiktionaries tragen - momentan momentan muss dieser noch manuell eingegeben werden
- **Französisch:** ...

Kasus	Singular	Plural
<b>Nominativ</b>	das Kraftfahrzeug	die Kraftfahrzeuge
<b>Genitiv</b>	des Kraftfahrzeugs	der Kraftfahrzeuge
<b>Dativ</b>	dem Kraftfahrzeug	den Kraftfahrzeugen
<b>Akkusativ</b>	das Kraftfahrzeug	die Kraftfahrzeuge

<http://de.wiktionary.org/w/index.php?title=Vorlage:Formatvorlage&oldid=13719> (Version vom 05.05.2004)



# Mikrostrukturprogramm 2005



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Form **Vorlagen nach Wortart** (iv) (Deutsch)

**Substantiv, m / f / n / u (?)**

**Silbentrennung:** Wort, Plural: Wör-ter

**Aussprache:**

IPA: [...], Plural: [...]

Hörbeispiele: —, Plural: —

**Bedeutungen:**

[1]

[2]

**Abkürzungen:**

**Herkunft:**

**Synonyme:**

[1]

**Gegenworte:**

**Oberbegriffe:**

**Unterbegriffe:**

**Beispiele:**

[1]

**Redewendungen:**

Format-  
änderungen

Kasus	Singular	Plural
<b>Nominativ</b>	{{{Wer oder was? (Einzahl)}}}	—
<b>Genitiv</b>	{{{Wessen? (Einzahl)}}}	—
<b>Dativ</b>	{{{Wem? (Einzahl)}}}	—
<b>Akkusativ</b>	{{{Wen? (Einzahl)}}}	—

Neue  
Angabeklassen

**Charakteristische Wortkombinationen:**

**Abgeleitete Begriffe:**

**Übersetzungen**

**Übersetzungen (Sprachen)**

- Arabisch: [] []]
- Chinesisch: [] []]
- **Englisch:** [1] [], [] []]
- Französisch: [] []]
- Russisch: [] []]
- Spanisch: [] []]

**Dialektausdrücke:**

- Alemannisch:
- Bairisch:
- Niedersächsisch:
- Ostmitteldeutsch:

? **Referenzen und weiterführende Informationen:**

[1] Wikipedia-Artikel „verschreiben“

**Ähnliche Wörter:**

# Mikrostrukturprogramm 2012



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

## Formatvorlage (Substantiv) (Deutsch)

### Substantiv, m / f / n / u (?)

#### Worttrennung:

Begriff, Plural: Begriffe

#### Aussprache:

IPA: [...], Plural: [...]

Hörbeispiele: —, Plural: —

#### Bedeutungen:

[1]

#### Abkürzungen:

[1]

#### Herkunft:

#### Synonyme:

[1]

#### Gegenwörter:

[1]

#### Oberbegriffe:

[1]

#### Unterbegriffe:

[1]

#### Beispiele:

[1]

#### Redewendungen:

Kasus	Singular	Plural
Nominativ	der Begriff	die Begriffe
Genitiv	des Begriffs	der Begriffe
Dativ	dem Begriff	den Begriffen
Akkusativ	den Begriff	die Begriffe

Terminologie

Zuordnung zu  
Bedeutungen

#### Charakteristische Wortkombinationen:

[1]

#### Abgeleitete Begriffe:

[1]

#### Übersetzungen

- Englisch: [1] [[]] → en
- Französisch: [1] [[]] → fr

- Italienisch: [1] [[]] → it
- Spanisch: [1] [[]] → es

#### Dialektausdrücke:

- Alemannisch:
- Bairisch:

- Kölsch:
- Ostmitteldeutsch:

Weitere  
Vorbelegungen

#### ? Referenzen und weiterführende Informationen:

- [1] Wikipedia-Artikel „Formatvorlage (Substantiv)“
- [1] Deutsches Wörterbuch von Jacob und Wilhelm Grimm „(Substantiv) Formatvorlage (Substantiv)“
- [1] Digitales Wörterbuch der deutschen Sprache „(Substantiv) Formatvorlage (Substantiv)“
- [1] canoo.net „(Substantiv) Formatvorlage (Substantiv)“
- [1] Uni Leipzig: *Wortschatz-Lexikon* „(Substantiv) Formatvorlage (Substantiv)“
- [1] The Free Dictionary „(Substantiv) Formatvorlage (Substantiv)“

#### Ähnliche Wörter:

[http://de.wiktionary.org/w/index.php?title=Vorlage:Formatvorlage\\_%28Substantiv%29&oldid=2172258](http://de.wiktionary.org/w/index.php?title=Vorlage:Formatvorlage_%28Substantiv%29&oldid=2172258) (Version vom 09.04.2012)

## 2. Datenbeschaffung

„[...] Referenzen, um den Lesern [...] die Gewissheit zu geben, dass die angegebenen **Bedeutungen** oder gar **das Wort selbst** nicht der Fantasie eines Einzelnen entsprungen sind und [...] eine weitere Verbreitung haben oder hatten. Dies zeigt sich daran, dass schon **andere Wörterbücher** entsprechende Angaben enthalten [...]“

### Lexikographische Ausgangsdatei in Wiktionary?

- Liste mit 47 Internet-Referenzen
  - ausschließlich sekundäre und tertiäre Quellen
- Liste mit 719 Literaturverweisen (+ wenige mehr auf Unterseiten)
  - sehr wenige primäre Quellen

# Quellenangaben in Wiktionary

Analyse der 96.339 deutschen Einträge in Wiktionary:

- $\approx 55.000$  (58%) Einträge mit sog. **Referenzen** (Gesamtartikel)
- $\approx 22.000$  (23%) Einträge mit sog. **Quellenangaben** (angabespez.)
- Insgesamt  $\approx 240.000$  **Verweise** auf  $< 27.000$  **eigenständige Quellen**

## Top 5 (bereits durch Vorlage vorgegeben)

- 1) 40.154 Wikipedia
- 2) 38.515 Wortschatz-Lexikon
- 3) 37.579 canoo.net
- 4) 33.596 DWDS
- 5) 12.741 TheFreeDictionary

# Quellenangaben in Wiktionary



- 6) 10.173 DWB
- 7) 6.693 Duden
- 8) 2.417 Duden Universalwörterbuch
- 9) 2.402 Duden: Großes Fremdwörterbuch
- 10) 1.706 Kluge: Etymologisches Wörterbuch
- 11) 1.514 Duden: Rechtschreibung
- 12) 1.005 Metzler Lexikon Sprache
- 13) 797 Pfeifer: Etymologisches Wörterbuch
- 14) 751 Duden: Herkunftswörterbuch
- 15) 741 Bußmann: Lexikon der Sprachwissenschaft
- 16) 710 Burkart: Vornamen
- 17) 698 OWID
- 18) 636 Duden: Vornamenlexikon
- 19) 532 Brockhaus
- 20) 453 Herder

# Primäre Quellen in Wiktionary

- Klarer Schwerpunkt auf sekundäre und tertiäre Quellen
- Keine einheitliche Liste
- Viele (primäre) Quellen sind nur einmal referenziert
- **Hauptsächlich Digitalisate und Online-abrufbare Quellen**
  - Zeitungsartikel und Online-Newsarchive
    - z.B. Spiegel, Focus, Zeit,...
  - Digitale Buchprojekte
    - Google Books
    - Projekt Gutenberg
- Wenig Belletristik

# Erarbeitung der Angaben

## Betreuungsgeld

### Betreuungsgeld (Deutsch) [Bearbeiten]

#### Substantiv, *n* [Bearbeiten]

##### Worttrennung:

Be-treu-ungs-geld, Plural: Be-treu-ungs-gel-der

##### Aussprache:

IPA: [bəˈtʁoʊŋsˌɡɛlt], Plural: [bəˈtʁoʊŋsˌɡɛldə]

Hörbeispiele: —, Plural: —

##### Bedeutungen:

[1] finanzielle, staatliche Unterstützung für Eltern, die ihr Kind zu Hause betreiben, anstatt es in eine Krippe zu schicken

##### Herkunft:

Determinativkompositum aus den Substantiven *Betreuung* und *Geld* mit dem Fugenelement *-s*

##### Beispiele:

[1] „Merkel ist indes fest entschlossen, den seit Monaten schwelenden Streit über das *Betreuungsgeld* jetzt zu beenden: Noch in diesem Monat soll ein Gesetz auf den Weg gebracht werden, dass Mütter und Väter, die ihre Kleinstkinder nicht in eine Krippe geben, mit zunächst 100 Euro im Monat unterstützt werden.“<sup>[1]</sup>

#### Übersetzungen [Bearbeiten]

- Englisch: [1] [[]] → en
- Französisch: [1] [[]] → fr

- Italienisch:
- Spanisch:

#### ? Referenzen und weiterführende Informationen:

- [1] Wikipedia-Artikel „*Betreuungsgeld*“
- [\*] Digitales Wörterbuch der deutschen Sprache „*Betreuungsgeld*“
- [1] Uni Leipzig: *Wortschatz-Lexikon* „*Betreuungsgeld*“

#### Quellen:

1. ↑ „Das ist ein K.-o.-Schlag“. In: *Welt Online*. 15. Mai 2012, ISSN 0173-8437 (URL [↗](#), abgerufen am 22. Mai 2012).

Kasus	Singular	Plural
Nominativ	das <i>Betreuungsgeld</i>	die <i>Betreuungsgelder</i>
Genitiv	des <i>Betreuungsgelds</i>	der <i>Betreuungsgelder</i>
Dativ	dem <i>Betreuungsgeld</i>	den <i>Betreuungsgeldern</i>
Akkusativ	das <i>Betreuungsgeld</i>	die <i>Betreuungsgelder</i>

Informationen sind nicht in referenzierten Artikeln zu finden  
**Neu erarbeitet oder fehlende Quellenangabe?**

Beispielsatz von Welt Online

**Quellenbasierte Erarbeitung lexikographischer Angaben oder Übernahme aus Sekundärquellen?**

# 3. Datenaufbereitung

- Keine einheitliche Wörterbuchbasis
- Keine systematische Exzerption erkennbar
  - weder manuell noch korpusgestützt
- Damit: keine lexikographische Zwischendatei
- **Phasen der Datenbeschaffung und Datenaufbereitung verschmelzen mit Datenauswertung**



„Schreibe Einträge! Das ist gar nicht so schwer! Weißt du **etwas, was viele andere nicht wissen**? Gibt es schon einen Beitrag über dein **Lieblingswort**? Über deinen **Beruf** oder **spezifische Fachwörter** [...]

Eine gute Idee ist es auch, mal in unsere automatisch erzeugte Liste der **gewünschten Seiten** zu schauen oder die Seite **fehlender Einträge**.“

- Keine systematische Auswahl
- Anlage von Einzelartikeln oder Artikelgruppen
  - alphabetische Teilstrecken; Wortlisten
  - bestimmte (Fach-)Themen
  - Benutzerwünsche
  - ...

# 4. Datenauswertung



## Erstellung von Wörterbuchartikeln

- Basierend auf Wortartenspezifischer Vorlage
- Lexikographische Instruktionen = Wiktionary-Hilfeseiten
- Nicht alle Artikelpositionen müssen ausgefüllt werden

## Überarbeiten von Wörterbuchartikeln

- Ergänzung
- Konkretisierung, Neuformulierung,...
- Fehlerkorrektur
- Belege/Quellen einfügen
- Irrelevante Information entfernen

# Überarbeitung von Wb-Artikeln

## Revision

- Stand des gesamten Artikels zu einem definierten Zeitpunkt
- Menge von Änderungsoperationen

## Änderungsoperation

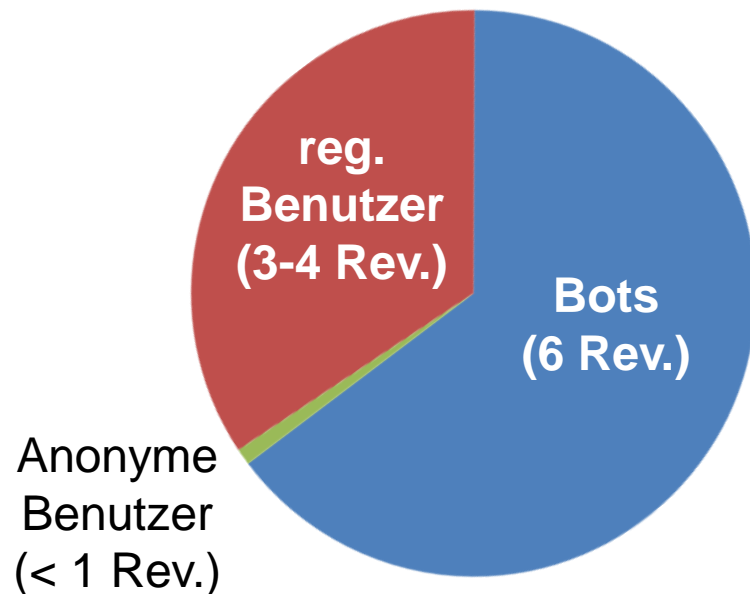
- **Ergänzung** (z.B. Ausspracheangabe)
- **Abwandlung** (z.B. Neuformulieren der Bedeutungsparaphrase)
- **Löschung** (z.B. Entfernen eines Beispielsatzes)
- **Formatierung** (Fettdruck, Klammerung,...)
- **Vandalismus & Zurücksetzen**
- Bezogen auf einzelne Artikelposition

Typologie in Anlehnung an Daxenberger&Gurevych (2012)

# Quantitative Analyse der Revisionen

Juli 2012:  $\approx 227.000$  Wb-Artikel mit  $\approx 2.400.000$  Revisionen

Durchschnittlich 10 Revisionen  
pro Artikel (Zipf-verteilt)



$\approx 838.000$  Revisionen von  
registrierten Benutzern, davon

- 70% substantielle Veränderungen (und 30% kleine Veränderungen)
- 75% mit Kommentar

ca. 113.000 Artikel  $> 1$  Revision

# Häufig geänderte Artikel

## Top 10

1) Haus	415x
2) Wasser	338x
3) gehen	303x
4) Katze	299x
5) Wort	296x
6) ich liebe dich	287x
7) ficken	283x
8) Hund	280x
9) ich	278x
10) Sprache	273x

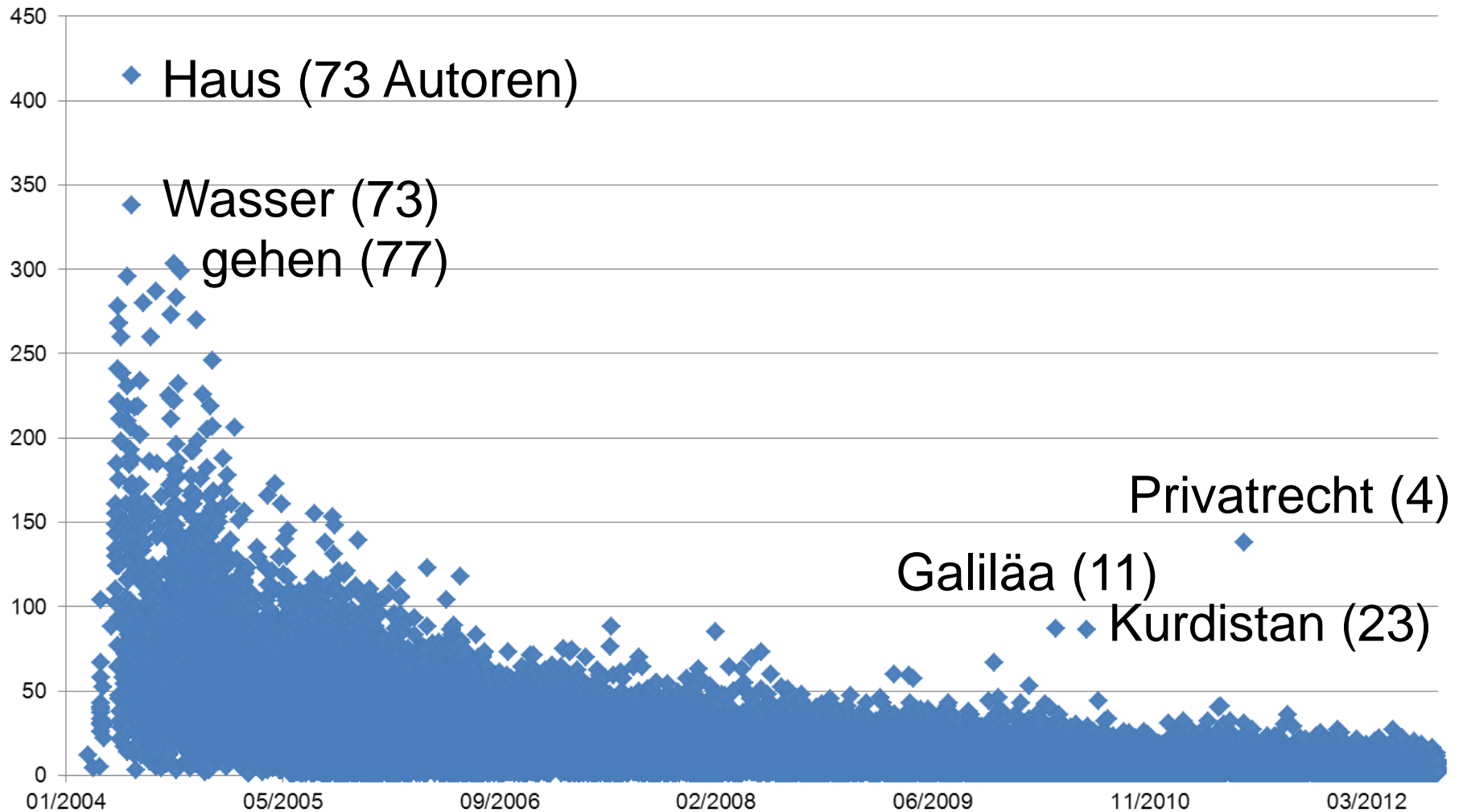
→ **Grundwortschatz**  
(> 50% der TOP 100  
in „GUT1 500“)

→ **Vulgäre Ausdrücke,  
Umgangssprache**

# Autoren pro Artikel

- Durchschnittlich **2–3 Autoren** pro Artikel (ohne Bots)
  - Artikel mit den meisten Autoren:
    - 1) gehen 77 Autoren
    - 2) Wasser 73 Autoren
    - 3) Haus 73 Autoren
    - 4) sein 70 Autoren
    - 5) Mensch 70 Autoren
- **wiederum vorw. Grundwortschatz (>64% in den TOP 100)**
- **Ältere Artikel haben tendenziell mehr Autoren**

# Revisionen nach Artikelalter



# „Hauptautoren“

Durchschnittlich 16% eines Artikels stammen von einem Autor

## **Benutzer #9200: themenbezogener Hauptautor**

- Hauptautor von 34 Artikeln,  $\emptyset = 38\%$  der Revisionen
  - Privatrecht, Sozialrecht, Gleichheitsrecht:  $>90\%$  der Revisionen
  - Sozialhilfe: 36% der Revisionen
  - Arbeitslosengeld: 9% der Revisionen (dichtauf mit „Zweitautor“)
- Insgesamt 584 Revisionen – 363 (62%) als Hauptautor

## **Benutzer #720: ergänzender Autor (hier: Übersetzungen)**

- Hauptautor von 8.891 Artikeln, aber: viele mit nur wenigen Änderungen
- Insgesamt 43.750 Revisionen – 20% als Hauptautor



# Kollaboration und Konflikte

ca. 3.400 Artikel mit zweitem Hauptautor  
(> 10% der Revisionen)

Artikel	Autor 1	Revisionen	Autor 2	Revisionen
Evolution	#720	38 (29%)	#6481	26 (20%)
Mohammed	#1812	24 (22%)	#28199	23 (21%)
Zigeuner	#15401	37 (27%)	#720	21 (15%)
backen	#28199	19 (12%)	#3869	18 (11%)
Schwarzäugige Susanne	#720	20 (40%)	#7294	17 (34%)
Airport	#1696	30 (47%)	#720	15 (23%)
...	...	...	...	...

**Qualitative Auswertung:** Hauptautoren arbeiten kooperativ oder zeitlich getrennt. Kein Beispiel für Kontroversen oder Konflikte.

## Treffpunkt [\[Bearbeiten\]](#)

Entschuldigt meine heftige Wortwahl, aber ich weiß nicht, was hier für "Fachleute" revertieren und sichten!!

- ein Ort, wo sich 2 Straßen treffen, nennt man "Straßenknick" oder "Straßenecke", im einfachsten Fall einfach nur Straße, wenn eine gerade verlaufende Straße von der a-Straße zur b-Straße wird.
- ein Ort, wo sich 3 Straßen treffen, wird Straßengabel oder auch Abzweigung genannt
- ein Ort, wo sich 4 oder mehr Straßen treffen, wird Kreuzung genannt. Das ist sinngleich mit der Aussage: ein Ort, wo sich 2 (oder mehr) Straßen kreuzen oder ein Ort, wo eine Straße eine zweite Straße quert.

Ich weiß wirklich nicht, was ihr hier den Lesern für Bedeutungen "unterjubeln" wollt, aber das ist schon grenzwertiges Verhalten. Da ich bei einer massiven Revertierung ohne jede Rückfrage erheblichen Sachverstand erwarte bzw. unterstelle, kann ich hier nur VM annehmen.

Ich möchte jetzt hier keinen Editwar anzetteln, weshalb ich die stattgefundene Revertierung meiner Änderungen nicht erneut ändere und bitte, jemand mit Sachkunde und Verständnis möge sich der Angelegenheit annehmen. —[JÄhh \(Diskussion\)](#) 12:03, 7. Mär. 2011 (MEZ)

Gerne. Eine Straße, die in eine Kreuzung führt, endet nicht dort, sondern läuft, salopp gesagt, einfach durch. So kommt man auf zwei oder mehr sich treffende Straßen. Definitorisch ebenso erfasst im [DWDS](#) [☞](#), in *Duden Das große Wörterbuch der deutschen Sprache in 10 Bänden. 3., völlig neu bearbeitete und erweiterte Auflage. Mannheim, Leipzig, Wien, Zürich: Dudenverlag 1999*, unter englisch "crossroads" im *Oxford Dictionary of English* und bei [Merriam-Webster](#) [☞](#). Wo steht es anders? Davon abgesehen ist dein Ton wieder einmal völlig unangemessen. —[Pill \(Kontakt\)](#) 12:50, 7. Mär. 2011 (MEZ)

Oder sehe ich jetzt das Problem nicht? Du siehst einen Unterschied zwischen "treffen" und "kreuzen"? —[Pill \(Kontakt\)](#) 14:44, 7. Mär. 2011 (MEZ)

Pill, ich kann es nicht fassen: Du definiert in Deinem Einleitungssatz eine Kreuzung damit, das Straßen, die in eine Kreuzung führen, dort nicht enden? Ist das jetzt hier pillepalle?

Wenn sich also 2 Straßen im Winkel von 45Grad treffen, dann ist das für die ganze Welt eine Spitzkehre oder auch eine sehr scharfe Kurve, aber für Euch oder Dich ist das eine Kreuzung - ja?

Wenn sich 3 Straßen jeweils im Winkel von 45 Grad treffen, dann ist das für Euch hier eine Kreuzung?

# 5. Satz- und Druckvorbereitung

- Erfolgt automatisch durch die Wiki-Software
- Markup-Sprache wird in fertig formatierte Wb-Artikel übersetzt

```
== Giraffe ({{Sprache|Deutsch}}) ==  
=== {{Wortart|Substantiv}}
```

[...]

```
{{Aussprache}}
```

```
:{{IPA}} {{Lautschrift|  
Aussprachewörterbuch|A=  
name="Mangold"/>, ''[[ö  
wissen.de|Wörterbuch|Gi  
{{Lautschrift|,ʒi'rafn}}
```

```
:{{Hörbeispiele}} {{feh
```

```
{{Bedeutungen}}
```

```
: [1] ''[[Zoologie]]:''  
[[mittelafrikanischen  
[[groß]]es, [[langbeini
```

## Giraffe

Giraffe (Deutsch) [Bearbeiten]

**Substantiv, f** [Bearbeiten]

**Worttrennung:**  
Gi-raf-fe, Plural: Gi-raf-fen

**Aussprache:**  
IPA: [ˌɡiˈrafə]<sup>[1]</sup>, auch: [ˌʒiˈrafə]<sup>[1]</sup>, *österreichisch* auch: [ˌʃiˈrafə]<sup>[2]</sup>; Plural: [ˌɡiˈrafn], auch: [ˌʒiˈrafn], *österreichisch*  
auch: [ˌʃiˈrafn]  
**Hörbeispiele:** —, Plural: —

- Bedeutungen:**
- [1] *Zoologie:* (in **afrikanischen Savannen** der **mittelafrikanischen Steppen** beheimatetes, in Herden lebendes) großes, **langbeiniges, pflanzenfressendes Säugetier** (*Giraffa camelopardalis*) mit einem 2-3 Meter **langen Hals**, dessen **kurzes Fell sandfarben gefärbt** und **unregelmäßig mit (rost-)braunen Flecken** versehen ist, dessen **Rücken stark abfällt** und auf dessen **Stirn** sich mehrere fellüberzogene **Knochenzapfen** befinden
  - [2] *Film, Jargon der Filmproduktion:* mit einem **langen, schwenkbaren Arm** versehenes **Gerät**, an dem beispielsweise eine **Kamera** oder ein (nicht sichtbares) **Mikrofon** montiert ist
  - [3] *Astronomie:* ein **Sternbild** am **nördlichen Sternhimmel**
  - [4] *umgangssprachlich:* hoch gewachsene, hagere **Person**
  - [5] *umgangssprachlich:* **langhalsige Person** (veraltet zumeist bezüglich Frauen)
  - [6] *umgangssprachlich:* **Persönlichkeit** von **hohem gesellschaftlichen Rang**
  - [7] *veraltet:* **hochmütige Person**
  - [8] *veraltet:* **Damenfrisur**, bei der das **Hinterhaar** in einer oder mehreren **großen Schleifen** auf dem **Scheitel** aufgesteckt wurde
  - [9] *pfälzisch veraltet:* **Person** mit **hervorstehenden Zähnen**

**Herkunft:**  
Ein seit dem 14. Jahrhundert<sup>[3]</sup> bezeugtes **Erbwort** aus dem **Mittelhochdeutschen**, das erstmals in der Form *schraffe* → *gmh* belegt ist<sup>[4]</sup>, die sich auf das **arabische** **جَرَاقَة** (DMG: *ǧurāfa*) →<sup>ar</sup>[4] beziehungsweise das **vulgärarabische** **جرافة** (DMG: *ǧrāfa*) →<sup>ar</sup> zurückführen lässt<sup>[5]</sup>. Die danach in Reiseberichten vorkommenden Formen, darunter das **spätmittelhochdeutsche** *seraffe* → *gmh* (14. Jahrhundert), das **frühneuhochdeutsche** *seraphen Pl.* (15. Jahrhundert),

Dieser Eintrag war 2012  
36. **Wort der Woche**.

Kasus	Singular	Plural
<b>Nominativ</b>	die Giraffe	die Giraffen
<b>Genitiv</b>	der Giraffe	der Giraffen
<b>Dativ</b>	der Giraffe	den Giraffen
<b>Akkusativ</b>	die Giraffe	die Giraffen



[1] kämpfende Giraffen im Ithala-Wildreservat in der südafrikanischen Provinz KwaZulu-Natal



[2] als Giraffe bezeichneter Kamerakran

# Der lexikographische Prozess



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

1. Vorbereitungsphase

Nachträgliche  
Änderungen?

2. Phase der Datenbeschaffung

3. Phase der Datenaufbereitung

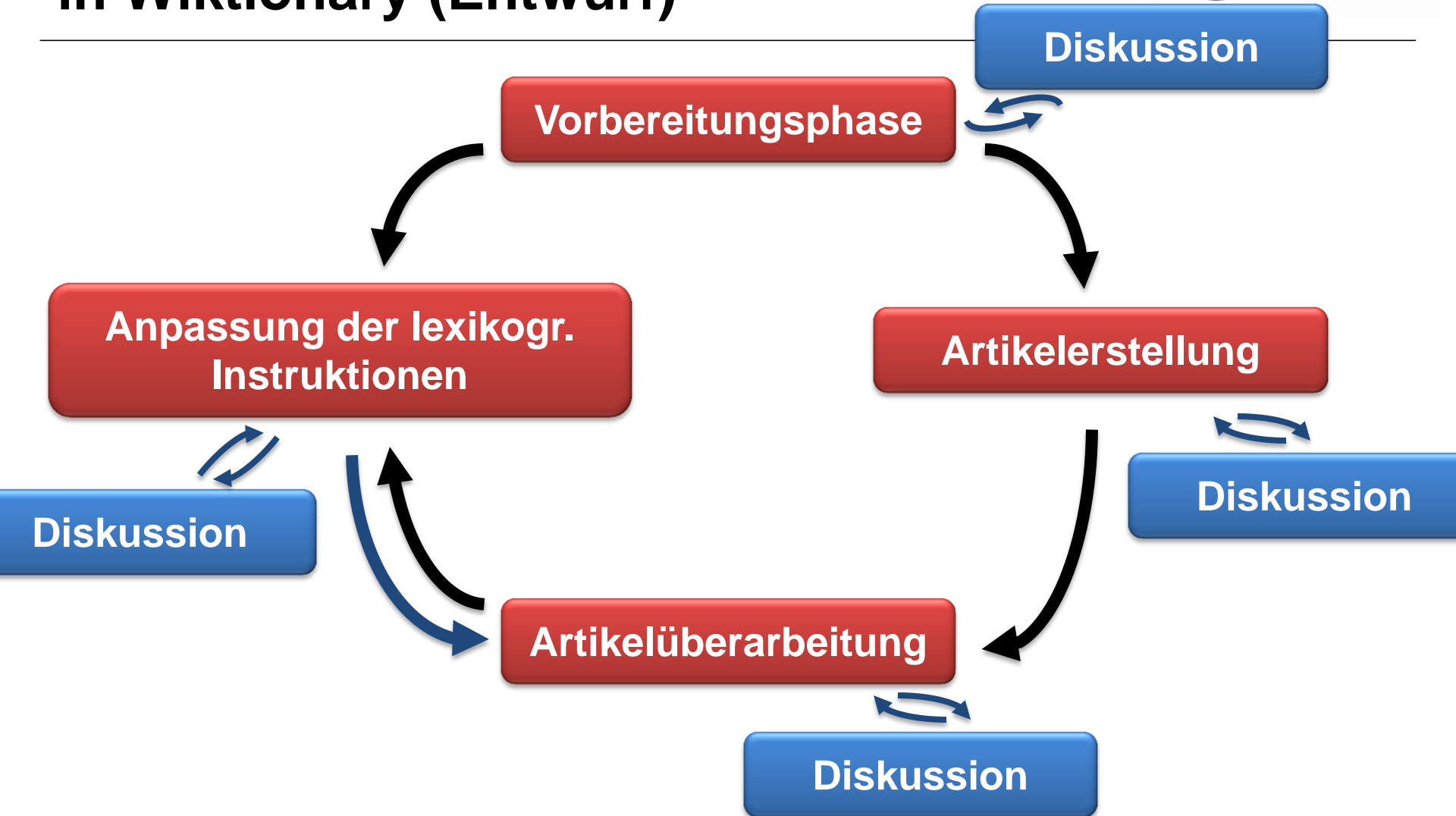
Phasen  
teilbar?

4. Phase der Datenauswertung

Überarbeitung  
von Artikeln?

5. Phase der Satz- und Druckvorbereitung

# Der lexikographische Prozess in Wiktionary (Entwurf)



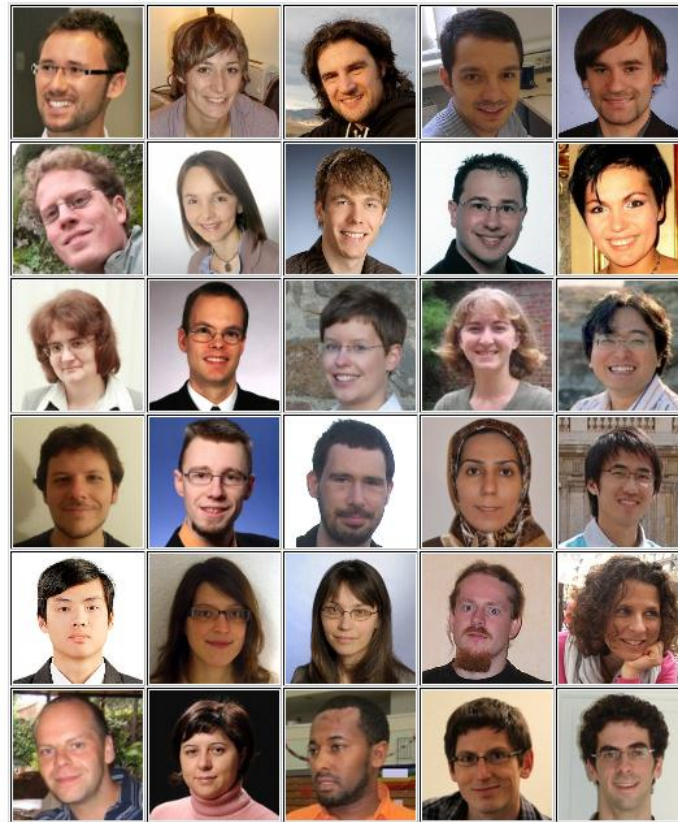
# Zusammenfassung

- **Wb-Konzeption** in Wiktionary erkennbar
  - Benutzerbezug bleibt unklar
- **Mikrostrukturprogramm** weitgehend stabil
- Keine **systematische Exzerption** erkennbar
- **Quellenangaben** von zentraler Bedeutung
  - Schwerpunkt auf sekundären Quellen
  - aber nicht ausschließlich
- Artikel haben 1–77 Autoren ( $\emptyset =$  **2–3 Autoren**)
  - Hauptautoren arbeiten eher **kooperativ**; wenig Konflikte

**Aus Print-Wb bekannte Phasen sind nur mäßig auf Wiktionary abbildbar; lexikographischer Prozess durch Überarbeitung und Diskussion getrieben**

# Vielen Dank für die Aufmerksamkeit!

## Ubiquitous Knowledge Processing



<http://www.ukp.tu-darmstadt.de>

Johannes Daxenberger & Iryna Gurevych: **A Corpus-Based Study of Edit Categories in Featured and Non-Featured Wikipedia Articles**, in: *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, Dezember 2012. Mumbai, Indien.

Oliver Ferschke, Torsten Zesch & Iryna Gurevych: **Wikipedia Revision Toolkit: Efficiently Accessing Wikipedia's Edit History**, in: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. System Demonstrations*, S. 97–102, Juni 2011. Portland, OR, USA.

Christian M. Meyer & Iryna Gurevych: **Wiktionary: a new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography**, Kapitel 13 in: *Electronic Lexicography*, S. 259–291, Oxford: Oxford University Press, November 2012.

Herbert Ernst Wiegand, Michael Beißwenger, Rufus H. Gouws, Matthias Kammerer, Angelika Storrer & Werner Wolski (Hrsg.): **Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research**, Band 1 (Systematische Einführung / Systematic Introduction, A–C), Berlin/New York: De Gruyter, 2010.








## Kontakt / Contact

### Christian M. Meyer

Technische Universität Darmstadt  
Ubiquitous Knowledge Processing Lab

 Hochschulstr. 10, 64289 Darmstadt, Germany

 +49 (0)6151 16-7477

 +49 (0)6151 16-5455

 meyer (at) ukp.informatik.tu-darmstadt.de

### Rechtliche Hinweise

Die Folien sind für den persönlichen Gebrauch der Vortragsteilnehmer gedacht. Im Vortrag verwendete Photographien, Illustrationen, Wort- und Bildmarken sind Eigentum der jeweiligen Rechteinhaber oder Lizenzgeber. Um Missverständnisse zu vermeiden, wäre eine kurze Kontaktaufnahme vor Weitergabe oder -nutzung der Vortragsmaterialien empfehlenswert. Sofern Sie Ihre Rechte verletzt sehen, bitte ich ebenfalls um Kontaktaufnahme zur Klärung der Sachlage.

### Legal Issues

The slides are intended for personal use by the audience of the talk. Photographies, illustrations, trademarks, or logos are property of the holder of rights. To avoid any misconceptions, I would strongly recommend to get in touch before reusing or redistributing the slides or any additional material of the talk. The same applies if you consider your rights infringed – please let me know to initiate further clarification.