

Qualität und Qualitätssicherung in nutzergenerierten Wörterbüchern? – Mögliche Stärken und Schwächen –

Andrea Abel und Christian M. Meyer



TECHNISCHE
UNIVERSITÄT
DARMSTADT

EURAC
research

Was ist ein gutes (Internet-)Wörterbuch? – Alte und neue Fragen zur Qualität lexikographischer Produkte im „digitalen Zeitalter“. EURALEX-Workshop des wissenschaftlichen Netzwerks „Internetlexikografie“, Bozen/Bolzano, Italien, 14. Juli 2014.

These 1: 2.000 Augen sehen mehr als 2
Kollaborative Qualitätssicherungsmaßnahmen
liefern einen deutlichen Mehrwert
für Online-Wörterbücher.

These 2: Viele Köche verderben den Brei
Eine zuverlässige Qualitätskontrolle ist nur
in redaktionell betreuten Wörterbüchern
möglich.

These 3: Dilltapp, bostitchen, ROFL, chambrieren, siissisoq

Nutzergenerierte Wörterbücher sind wertvolle Ressourcen für Sprachvarietäten und Mehrsprachiges.

These 4: Bleiben Ehrfurcht, Podeste oder Favoriten auf der Strecke?

Nutzergenerierte Wörterbücher weisen beachtliche Lücken auf.

These 5: Auf den Schultern vieler

Durch die systematische Einbindung und Steuerung von Nutzer_innen wird ein beschleunigter Wörterbuchausbau möglich.

These 6: Quick and dirty?

Nutzerbeteiligung beschleunigt einen qualitätsvollen Ausbau von Wörterbüchern nicht.

These 7: Laienlexikograph_innen sind die besseren Benutzer_innen

Mitmachwörterbücher erfüllen einen Didaktikzweck und begünstigen die erfolgreiche Wörterbuchbenutzung.

These 8: Und es braucht sie doch, die Lexikograph_innen!

Laienlexikograph_innen orientieren sich an Produkten der „Qualitätslexikographie“ und sind konservativ.

These 9: Ein Schatz für die Sprachtechnologie

Offen lizenzierte Wörterbücher bilden eine wichtige Grundlage zur automatischen Verarbeitung von Sprache.

These 10: Kopieren, kopieren, kopieren
Nutzergenerierte Wörterbücher sind ein Schatz für die Sprachtechnologie, vielfach handelt es sich aber um Schätze anderer.



Dieser Eintrag oder Abschnitt bedarf einer Überarbeitung. Hilf bitte mit, ihn zu verbessern, und entferne anschließend diese Markierung.

Folgendes ist zu überarbeiten: Bedeutungsdefinitionen so umformulieren, dass es keine Synonymlisten sind

aufrecht (Deutsch)

Adjektiv

Worttrennung:

auf·recht, Komparativ:
auf·rech·ter, Superlativ: am
auf·rech·tes·ten

Positiv

aufrecht

Komparativ

aufrechter

Superlativ

am aufrechtsten

Alle weiteren Formen: **aufrecht (Deklination)**

Bedeutungen:

- [1] hochgerichtet, gerade
- [2] aufrichtig, unverfälscht, authentisch
- [3] rechtschaffen, redlich

Abb. 1: Baustellenkennzeichnungen in *Wiktionary*
<http://de.wiktionary.org/wiki/aufrecht> (07.01.2014)

Is this translation correct and useful?

English	loodering [Scot.] [Irish]
German	heftige Prügelei {f}
Class	noun
Comment	<ul style="list-style-type: none">• http://www.oed.com/view/Entry/... " 'loodering n. a severe beating; a hiding. 1866 W. Gregor Dial. Banffshire (Philol. Soc.) 227 Llouthera severe beating. 1912 J. Campbell Judgment i. 9 She'll get over it. It's not the 1985 L. Shannon in Stories (Boston) No. 12 14 Whenever s biggest loodering she's ever had in her whole life. "

YES (100% correct) | NO / MAYBE ...

Abb. 2: Kollaborative Qualitätsbewertung in *dict.cc*
<http://contribute.dict.cc/?action=wizard> (27.01.2014)

admit (*third-person singular simple present **admits**, present participle **admitting**, simple past and past participle **admitted***)

1. (*transitive*) To **allow** to **enter**; to grant **entrance**, whether into a place, or into the mind, or **consideration**; to **receive**; to take.

*A ticket **admits** one into a playhouse.*

*They were **admitted** into his house.*

*to **admit** a serious thought into the mind*

*to **admit** evidence in the trial of a cause*

2. (*transitive*) To allow (one) to enter on an office or to enjoy a privilege; to **recognize** as qualified for a **franchise**.

*to **admit** an attorney to practice law*

*the prisoner was **admitted** to bail*

Abb. 3: admit im englischen *Wiktionary*
<http://en.wiktionary.org/wiki/admit> (30.06.2014)

Heinrich Tischner
Fehlheimer Straße 63
64625 Bensheim

Südhessisch Diltapp

Dill-tapp m. 'Diltapp'

- (0) 'unbeholfener Tritt auf die Diele'
 - 'Mensch, der unbeholfen geht'
 - 'Tollpatsch, ungeschickter Mensch' **EErf**
 - 'ein Fabeltier, das ein Uneingeweihter fangen soll'

nach oben
Übersicht

ShWb 1,1526 | Synonyme **Fabelwesen**

Abb. 4: Diltapp in der *Datenbank Südhessisch*

<http://www.heinrich-tischner.de/22-sp/6mda/sh/wort/d/di/dil/d-tapp.htm> (25.06.2014)

Bedeutungen:

[1] *schweizerisch*: mit u-förmigen Metallklammern heften, tackern

Herkunft:

Das Verb *bostitchen* wurde zu dem Substantiv *Bostitch* gebildet.^[1]

Dieses wiederum leitet sich von dem 1896 gegründeten Unternehmen *Boston Wire Stitcher Company* ab, das Drahtheftmaschinen herstellte, die zum Heften etwa von Büchern oder Kalendern dienten.^[2] 1948 verkürzte man den Namen zu *Bostitch*.^[2] Dieser hat sich in der Schweiz als übliche Bezeichnung für Tacker durchgesetzt.^[2]

Synonyme:

[1] in Deutschland: tackern

Beispiele:

[1] Im Werkunterricht wird gehämmert, gesägt und *gebostitcht*.

[1] Zum Bespannen biegen Sie die Pappe etwas, legen Sie den Stoff auf und *bostitchen* Sie die Enden rundherum fest.^[3]

Abb. 5: *bostitchen* im deutschen *Wiktionary*
<http://de.wiktionary.org/wiki/bostitchen> (01.06.2014)

Verb [Bearbeiten]		
	Person	Wortform
Worttrennung: cham·brie·ren, Präteritum: cham·brier·te, Partizip II: cham·briert	Präsens	ich chambriere
		du chambrierst
		er, sie, es chambriert
Aussprache: IPA: [ʃä'br̩:ɾən], Präteritum: [ʃä'br̩:ɾtə], Partizip II: [ʃä'br̩:ɾt]	Präteritum	ich chambrierte
	Konjunktiv II	ich chambrierte
Hörbeispiele: —, Präteritum: —, Partizip II: — Reime: -i:ɾən	Imperativ	Singular chambriere!
		Plural chambriert!
Bedeutungen: [1] <i>Weinbau</i> : das allmähliche Erwärmen von Rotwein von Keller- auf Zimmertemperatur	Perfekt	Partizip II
		chambriert Hilfsverb haben
Beispiele: [1] Kannst du bitte den Rotwein <i>chambrieren</i> ?		
Alle weiteren Formen: <i>chambrieren</i> (Konjugation)		

Abb. 6: chambrieren im deutschen *Wiktionary*
<http://de.wiktionary.org/wiki/chambrieren> (06.02.2014)

1. ROFL



ROFL is an internet acronym for Rolling On Floor Laughing, and like all things internet it has adapted to nerd culture and has taken on new real-life use, though usually in a satirical way.

ROFL can be merged with other acronyms to create super l33t acronyms that "r0x0r y0ur b0x0rs" per se. Some examples of the merged acronyms are as follows:

Abb. 7: ROFL im *Urban Dictionary*

<http://www.urbandictionary.com/define.php?term=ROFL> (25.06.2014)

• Afrikaans: [1] renoster → af	• Kroatisch: [1] nosorog → hr
• Amharisch: [1] አቅራቢ (äqeraris) → am	• Kurdisch: [1] Yek Qoç → ku
• Arabisch: [1] وحيد القرن (w?ḥīd al-q?r?n) → ar	• Latein: [1] rhinoceros → la
• Armenisch: [1] ռնգեղջյուր (rrngeghjhyowr) → hy	• Litauisch: [1] raganosis → lt
• Bengalisch: [1] গন্ডার (gandār) → bn	• Malaiisch: [1] badak → ms
• Bosnisch: [1] nosorog → bs	• Marathi: [1] गेंडा (geṇḍā) → mr
• Bulgarisch: [1] носорог (nosorog) → bg	• Niederländisch: [1] neushoorn → nl
• Chinesisch: [1] 犀牛 (xīniú) → zh	• Norwegisch: [1] neshorn → no
• Dänisch: [1] næsehorn → da	• Okzitanisch: [1] rinocèros → oc
• Englisch: [1] rhinoceros → en, rhino → en	• Persisch: [1] کَرگدن (kargadan) → fa
• Esperanto: [1] rinocero → eo	• Polnisch: [1] nosorożec → pl
• Finnisch: [1] sarvikuono → fi	• Portugiesisch: [1] rinoceronte → pt
• Frankoprovenzalisch: [1] rinocèros → frp	• Rumänisch: [1] rinocer → ro
• Französisch: [1] rhinocéros → fr	• Russisch: [1] носорог (nosorog) → ru
• Galicisch: [1] rinoceronte → gl	• Schwedisch: [1] noshörning → sv
• Georgisch: [1] მარტორკა (mart'orka) → ka	• Serbisch: [1] носорог → sr / nosorog
• Griechisch (Neu-): [1] ρινόκερος (rinókeros) → el	• Slowenisch: [1] nosorog → sl
• Grönländisch: [1] siissisoq → kl	• Niedersorbisch: [1] nosorogak → dsb

Abb. 8: Wiktionary-Übersetzungen für Nashorn
<http://de.wiktionary.org/wiki/Nashorn> (13.05.2014)

▼ Nashorn : A large herbivorous pachyderm of the family Rhinocerotidae

▼ Definition

Sprache	Text
Englisch	A large herbivorous pachyderm of the family Rhinocerotidae, with thick, gray skin and one or two horns on their snouts.
Französisch	Grand pachyderme herbivore de la famille des Rhinocerotidae, avec une épaisse peau grise et une ou deux cornes sur leur nez.
Italienisch	Grande pachiderma erbivoro della famiglia Rhinocerotidae, con una spessa pelle grigia e una o due corna sui loro musi.
Niederländisch	Een grote planteneterende pachyderm uit de familie Rhinocerotidae, met een dikke, grijze huid en een of twee hoorns op hun snuit.
Portugiesisch	Um grande paquiderme herbívoro da família dos Rhinocerotidae, de pele grossa e cinzenta, e um ou dois cornos sobre o nariz.

Abb. 9: Bedeutungsparaphrasen für Nashorn in OmegaWiki
<http://www.omegawiki.org/Expression:Nashorn> (25.06.2014)

- Effekthascherei
- Effizienzklasse
- Egregor
- Ehehindernis
- Ehemodell
- Ehrendoktor
- Ehrendoktorhut
- Ehrensache
- Ehrerbietung
- Ehrfurcht
- Ehrung
- Plug-in-Hybrid
- Pneu
- Podest
- Poikilodermie
- polare Emergenz
- Politikredaktion
- Politologe
- Polizeiauto
- Polizeigewalt
- Polsterreiniger
- Polychondritis

Abb. 10: Ausschnitt der *Wiktionary*-Wunschliste

<http://de.wiktionary.org/wiki/Wiktionary:Wunschliste> (30.06.2014)

Substantiv, *m* [Bearbeiten]

Worttrennung:

Fa·vo·rit, Plural: Fa·vo·ri·ten

Aussprache:

IPA: [favo'ri:t], Plural:

[favo'ri:tən]

Hörbeispiele:

Reime: -i:t

Kasus	Singular	Plural
Nominativ	der Favorit	die Favoriten
Genitiv	des Favoriten	der Favoriten
Dativ	dem Favoriten	den Favoriten
Akkusativ	den Favoriten	die Favoriten

Bedeutungen:

[1] Lebewesen, das von jemandem bevorzugt wird

[2] *Sport*: Teilnehmer mit der größten Aussicht, in einem Vergleich zu gewinnen

Herkunft:

vom lateinischen *favere* ^{→ *la*} für „begünstigen“;^[1] zu *favor* ^{→ *la*} wörtlich für „Wohlwollen“^{[2][3]}

Abb. 11: Favorit im deutschen *Wiktionary*
<http://de.wiktionary.org/wiki/Favorit> (30.06.2014)

Favorit, der

Wortart: Substantiv, maskulin

Häufigkeit: ■■■■

Rechtschreibung ⓘ

Worttrennung:

Fa|vo|rit

Bedeutungen ⓘ

- a.** jemand, der bevorzugt, anderen vorgezogen wird; begünstigte Person, Liebling
b. (veraltet) Günstling, Geliebter
- Teilnehmer an einem Wettbewerb mit den größten Aussichten auf den Sieg

Abb. 12: Favorit in *Duden online*

<http://www.duden.de/rechtschreibung/Favorit> (30.06.2014)

Favorit Aussprache: ▶
mask., -en, -en
Herkunft: Französisch

1 **Günstling, Liebling**
er wurde der Favorit des Kaisers +
etw. Bevorzugtes, Beliebtes
großzügige, bunte Drucke, spitze Schuhe waren Favoriten der Mode

2 **Sport Sportler, Mannschaft, besonders auch Rennpferd mit den größten Gewinnchancen**
die Mannschaft ging als Favorit ins Spiel +
übertragen
er ist Favorit bei der Präsidentschaftswahl

Abb. 13: Favorit im *DWDS-Wörterbuch*
<http://www.dwds.de/?qu=Favorit> (30.06.2014)

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).

Abb. 14: Quellenangaben in *Wiktionary*
<http://de.wiktionary.org/wiki/Betreuungsgeld> (28.09.2012)

-
- [1] Abel Andrea/Annette Klosa: Ihr Beitrag bitte! – Der Nutzerbeitrag im Wörterbuchprozess, in Andrea Abel/Annette Klosa (Hg.): *Der Nutzerbeitrag im Wörterbuchprozess. 3. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“* (= OPAL – Online publizierte Arbeiten zur Linguistik 2/2014), S. 3-7, Mannheim: Institut für Deutsche Sprache, 2014.
<http://pub.ids-mannheim.de/laufend/opal/pdf/opal2014-4.pdf>
- [2] Andrea Abel/Christian M. Meyer: The dynamics outside the paper: user contributions to online dictionaries, in: *Proceedings of the 3rd eLex conference 'Electronic lexicography in the 21st century: thinking outside the paper'*, S. 179–194, 2013. Tallinn, Estland.
http://eki.ee/elex2013/proceedings/eLex2013_13_Abel+Meyer.pdf
- [3] Alice Yin Wa Chan/Andrew Taylor: Evaluating Learner Dictionaries: What the Reviews Say. In: *International Journal of Lexicography*, 14, S. 163–180, 2001.
- [4] Pedro A. Fuertes-Olivera: The Function Theory of Lexicography and Electronic Dictionaries: Wiktionary as a Prototype of Collective Free Multiple-Language Internet Dictionary, in Henning Bergenholtz/Sandro Nielsen/Sven Tarp (Hg.): *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow* (=Linguistic Insights: Studies in Language and Communication 90), S. 99–134, Bern: Peter Lang, 2009.
- [5] Patrick Hanks: Word Meaning and Word Use: Corpus evidence and electronic lexicography, in Sylviane Granger/Magali Paquot (Hg.): *Electronic Lexicography*, S. 57–82, Oxford: Oxford University Press, 2012.
- [6] Franz Josef Hausmann: Dictionary Criminality, in Franz J. Hausmann/Oskar Reichmann/Herbert E. Wiegand/Ladislav Zgusta/Rufus Gouws (Hg.): *Wörterbücher: Ein internationales Handbuch zur Lexikographie* (= Handbücher zur Sprach- und Kommunikationswissenschaft 5.1), S. 97–101. Berlin/New York: de Gruyter, 1989.

- [7] Sidney I. Landau: *Dictionaries: The art and craft of lexicography*. 2nd edition. Cambridge: Cambridge University Press, 2001.
- [8] Robert Lew: User-generated content (UGC) in English online dictionaries, in Andrea Abel/Annette Klosa (Hg.): *Der Nutzerbeitrag im Wörterbuchprozess. 3. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“* (= OPAL – Online publizierte Arbeiten zur Linguistik 2/2014), S. 8–26, Mannheim: Institut für Deutsche Sprache, 2014.
<http://pub.ids-mannheim.de/laufend/opal/pdf/opal2014-4.pdf>
- [9] Michael Matuschek/Christian M. Meyer/Iryna Gurevych: Multilingual Knowledge in Aligned Wiktionary and OmegaWiki for Translation Applications, *Translation: Computation, Corpora, Cognition – Special Issue “Language Technology for a Multilingual Europe”* 3(1):87–118, 2013.
<http://www.t-c3.org/index.php/t-c3/article/view/20>
- [10] Christian M. Meyer/Iryna Gurevych: Wiktionary: A new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography, in Sylviane Granger/Magali Paquot (Hg.): *Electronic Lexicography*, S. 259–291, Oxford: Oxford University Press, 2012.
- [11] Christian M. Meyer: *Wiktionary: The Metalexigraphic and the Natural Language Processing Perspective* (= tprints 3654). Dissertation, Darmstadt: Technische Universität Darmstadt 2013.
<http://tprints.ulb.tu-darmstadt.de/3654/>
- [12] Christian M. Meyer/Iryna Gurevych: Methoden bei kollaborativen Wörterbüchern, *Lexicographica* 30, erscheint 2014.
- [13] Carolin Müller-Spitzer (Hg.): *Using Online Dictionaries* (= *Lexicographica: Series maior. Supplementbände zum Internationalen Jahrbuch für Lexikographie. Supplementary volumes to the international annual for lexicography, Band 145*). Berlin/Boston: de Gruyter, 2014.

- [13] Carolin Müller-Spitzer (Hg.): *Using Online Dictionaries* (= Lexicographica: Series maior. Supplementbände zum Internationalen Jahrbuch für Lexikographie. Supplementary volumes to the international annual for lexicography, Band 145). Berlin/Boston: de Gruyter, 2014.
- [14] Sheizaf Rafaeli/Yaron Ariel: Online Motivational Factors: Incentives for Participation and Contribution in Wikipedia, in Azy Barak (Hg.): *Psychological Aspects of Cyberspace: Theory, Research, Applications*, S. 234–267, Cambridge: Cambridge University Press, 2008.
- [15] Michael Rundell: ‘It works in practice but will it work in theory?’ The uneasy relationship between lexicography and matters theoretical, in Ruth Vatdedt Fjeld/Julie Matilde Torjusen (Hg.): *Proceedings of the 15th EURALEX International Congress*, S. 47–92, 2012, Oslo, Norwegen.
- [16] Katrin Thier: Das Oxford English Dictionary und seine Nutzer, in Andrea Abel/Annette Klosa (Hg.): *Der Nutzerbeitrag im Wörterbuchprozess. 3. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“* (= OPAL – Online publizierte Arbeiten zur Linguistik 2/2014), S. 62–69, Mannheim: Institut für Deutsche Sprache, 2014.
<http://pub.ids-mannheim.de/laufend/opal/pdf/opal2014-4.pdf>
- [17] Torsten Zesch/Iryna Gurevych: Wisdom of Crowds versus Wisdom of Linguists – Measuring the Semantic Relatedness of Words, *Journal of Natural Language Engineering* 16(1):25–59, 2010.
- [18] https://meta.wikimedia.org/wiki/Grants:IEG/Pronunciation_Recording_%28Finish_incomplete_GSoC_project%29 (25.06.2014)
- [19] https://meta.wikimedia.org/wiki/Wikimedia_News#April_2014 (25.06.2014)
- [20] <http://stats.wikimedia.org/wiktionary/EN/TablesArticlesTotal.htm> (25.06.2014)
- [21] <http://wikimedia.de/wiki/Schulprojekt> (25.06.2014)

Kontakt / Contact

Andrea Abel

European Academy of Bozen/Bolzano (EURAC)
Institute for Specialised Communication
and Multilingualism

 Drususallee 1, 39100 Bolzano/Bozen, Italy


 +39 0471 055 121


 +39 0471 055 199


 andrea.abel (at) eurac.edu

Christian M. Meyer

Technische Universität Darmstadt
Ubiquitous Knowledge Processing Lab

 Hochschulstr. 10, 64289 Darmstadt, Germany

 +49 (0)6151 16–5386

 +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.