



# Community Proofreading

Zielstellung, Implementation, Evaluation

Sebastian Nordhoff  
October 2, 2019  
Language Science Press

- › im Open Access geht es meistens um Zugang, d.h. ums Lesen
- › Open Publishing geht darüber hinaus:
  - › Open source platforms
  - › Open formats
  - › Open protocols
  - › Open bookkeeping
  - › Open peer review
  - › **Community proofreading**

- › Forschende nehmen je nach Uhrzeit ganz verschiedene Rollen ein
  - › Autor, Gutachterin, Leserin
- › je jünger desto Leser
- › je älter desto Gutachter
- › komplexes Ökosystem
- › Integration aller Ebenen im community-based Publishing

- › eingekaufte Dienstleistung
- › 1 Person
- › spezialisiert in Stil und Richtlinien
- › fachlicher Hintergrund möglich aber nicht zwingend
- › normalerweise kein Spezialwissen in spezifischen Teilgebieten der Linguistik



- › Open-Access-Verlag in der Sprachwissenschaft
- › 100+ Bücher seit 2014
- › 350 Community Proofreader

<p><b>Natural causes of language: Frames, biases, and cultural transmission</b> N. Enfield (Author)</p>	<p><b>A topology of marked languages</b> Cornelia Fuchs (Author)</p>	<p><b>The Talking Heads experiment: Origins of words and meanings</b> Luc L. Lévesque (Author)</p>	<p><b>Grammaticalization in the North: Noun phrase morphemes in Scandinavian vernaculars</b> Oleksa Dahl (Author)</p>	<p><b>A grammar of Fijian Samsi</b> Joshua Wilbur (Author)</p>	<p><b>Prosodic detail in Neogutierrezian</b> Francesco Cengini (Author)</p>
<p><b>Linguistic variation, identity construction and cognition</b> Karin K. Dräger (Author)</p>	<p><b>Language strategies for the domain of labor</b> Jens Börs (Author)</p>	<p><b>A grammar of Ekele</b> Doro Schackner (Author)</p>	<p><b>Thoughts on grammaticalization</b> Christoph Lehmann (Author)</p>	<p><b>How mobile robots can self-organize a vocabulary</b> Paul Gahr (Author)</p>	<p><b>A grammar of Mawé</b> Lisa Berglund (Author)</p>
<p><b>Roots of language</b> Derek Bickerton (Author)</p>	<p><b>The empirical base of linguistics: Grammatical judgments and linguistic methodology</b> Carsten T. Schuler (Author)</p>	<p><b>A grammar of Palula</b> Henk Lippgen (Author)</p>	<p><b>New directions in corpus-based translation studies</b> Claudia Felton, Frederic Zeman (Volume Editor)</p>	<p><b>Grammatical theory: From transformational grammar to construction-based approaches</b> Stefan Müller (Author)</p>	<p><b>Advances in the study of Slavic languages and linguistics</b> Catherine Ruy, Björn Jansen (Volume Editor)</p>
<p><b>The evolution of grounded spatial language</b> Michael Spranger (Author)</p>	<p><b>The evolution of case grammar</b> Bernard Trapp (Author)</p>	<p><b>The future of dialects: Selected papers from Methods in Dialectology XV</b> Marie-Hélène Cobb, Bernice Knoch, John Nerbonne (Volume Editor)</p>	<p><b>Syntax und Valenz: Der Modellierung von Verbstrukturauswahl und -interpretation</b> Svenja Lohr (Author)</p>	<p><b>Adjective attribution</b> Michael Heller (Author)</p>	<p><b>A grammar of Kanyok</b> Anna Steinhilber (Author)</p>
<p><b>Sprachliche Sozialisation: Jüdischkeit in der deutschsprachigen Literatur (18.-20. Jahrhundert)</b> Lisa Schäfer (Author)</p>	<p><b>Forthcoming: Language technologies for a multilingual Europe</b> Gergely Bencsik, Daniel Gábor, Zoltán Székely, András Vicsi (Volume Editor)</p>	<p><b>Eye-tracking and Applied Linguistics</b> Silvia Hansen Schütz, Sandra Grunz (Volume Editor)</p>	<p><b>Forthcoming: Grammaticalised constructions in languages: Une étude comparative des phrases sans verbe en roumain et en français</b> Gabriela Stăle (Author)</p>	<p><b>Einführung in die grammatische Beschreibung der Deutschen: Zweite, überarbeitete Auflage</b> Roland Schüller (Author)</p>	<p><b>New perspectives on calico: conference implications for translation</b> Karin Mennel, Dorothea Luginbuhl, Barbara Krenz (Volume Editor)</p>
<p><b>Time in Yenching: Nal lexical bases and morphosyntax</b> Alexa Michael (Author)</p>	<p><b>A grammar of Kapa Nui</b> Paula Newman (Author)</p>	<p><b>Forthcoming: Attribution constructions in North-Eastern Neo-Aramaic</b> Jens Gutsch (Author)</p>	<p><b>Forthcoming: Empirical modelling of translation and interpreting</b> Silvia Hansen Schütz, Oliver Gals, Sascha Hoffmann, Bernd Meyer (Volume Editor)</p>	<p><b>Forthcoming: Reflections on Bibliolinguistics</b> Gerd Bock (Author)</p>	<p><b>Diversity in African languages: Selected papers from the 46th Conference on African Linguistics</b> Doris L. Payne, Sara Plet Hout, Beate Veldman (Volume Editor)</p>
<p><b>Annotation, exploitation and evaluation of parallel corpora: TC3</b> Silvia Hansen Schütz, Sandra Grunz, Oliver Gals (Volume Editor)</p>	<p><b>Forthcoming: Beiträge zur deutschen Grammatik</b> Tina K. Hilde (Author), Stefan Müller, Frank Richter, Margarete (Volume Editor)</p>	<p><b>The Zingales Dictionary and grammar sketch</b> Toni Schick (Author)</p>	<p><b>A grammar of Pagan Malay</b> Angela Rieger (Author)</p>	<p><b>Forthcoming: Unity and diversity in grammaticalization scenarios</b> Walter Stroh, Achim Machobane (Volume Editor)</p>	<p><b>Tonal placement in Tachikawa</b> Toshiyuki Kato (Author)</p>
<p><b>Forthcoming: Order and structure in syntax II: Subordinate and argument structure</b> Laura A. Bailey, Michelle Sheehan (Volume Editor)</p>	<p><b>Forthcoming: Order and structure in syntax I: Word order and syntactic structure</b> Laura A. Bailey, Michelle Sheehan (Volume Editor)</p>	<p><b>Forthcoming: A equívoco da língua materna em português: Questões gerais e específicas do Português</b> Marta João Freitas, Ana Lúcia de Santos (Volume Editor)</p>	<p><b>Forthcoming: Further investigations in the syntax of infinitives</b> Cristina Tiberius (Volume Editor)</p>	<p><b>Dependencies in language: On the central argument of linguistic systems</b> N. J. Enfield (Volume Editor)</p>	<p><b>On looking towards (and beyond) the future: A collection of papers from the 46th Conference on African Linguistics</b> Doris L. Payne, Sara Plet Hout, Beate Veldman (Volume Editor)</p>

# Community proofreading



- › Crowdsourcing
- › freiwillig
- › viele Korrektoren, häufig jünger
- › sehr häufig spezialisiert im Teilgebiet
- › intrinsisches Interesse
- › relativ gesehen geringere Expertise in Stilistik und Richtlinien



- › Proofreading Queue: neuer Titel alle 2 Wochen
- › Titel wird montags angekündigt
- › Interessierte können sich für ein Kapitel melden
- › Kapitel werden mittwochs zugewiesen
- › Dauer: 4 Wochen
- › Software: Paperhive



Document

Discussions 317

Activity

which would yield only default agreement on the RC predicate, contrary to fact.

## 2.2 The Genitive of Quantification phenomenon

The Genitive of Quantification phenomenon has been described to a large extent for Slavic languages in Bošković (2006); Franks (1994; 2002); Przepiórkowski (2004); Rutkowski (2002); and Willim (2003), to name but a few. In Polish, genitive case marking is forced on a noun which is modified by a higher numeral or a lower virile numeral, as well as by certain quantifiers such as *wiele* 'many', *kilka* 'a few', *para* 'a couple of', etc. Such numeral phrases do not induce subject-verb agreement in main clauses, as can be seen in (17), in which the verb obligatorily appears in the 3SG neuter form, regardless of the grammatical gender of the noun.

- (17) a. Siedmiu mężczyzn weszło/\*weszli do domu.  
seven.ACC men.GEN,VIR entered.3SG,NEUT/3PL,VIR into house  
'Seven men entered the house.'
- b. Siedem kobiet weszło/\*weszły do domu.  
seven.ACC women.GEN, NON-VIR entered.3SG,NEUT/\*3PL,NON-VIR into house  
'Seven women entered the house.'

The analysis of Polish GoQ structures proposed in Witkoś & Dziubała-Szrejberska (2016) follows the idea that probing for phi-features is possible for T

x



Alexandr Rosen · 18 days ago

### N instead of NEUT

according to

<https://www.eva.mpg.de/lingua/resources/glossing-rules.php>

Show more



Alexandr Rosen · 18 days ago

### GEN.VIR

period as the separator  
according to

<https://www.eva.mpg.de/lingua/resources/glossing-rules.php>

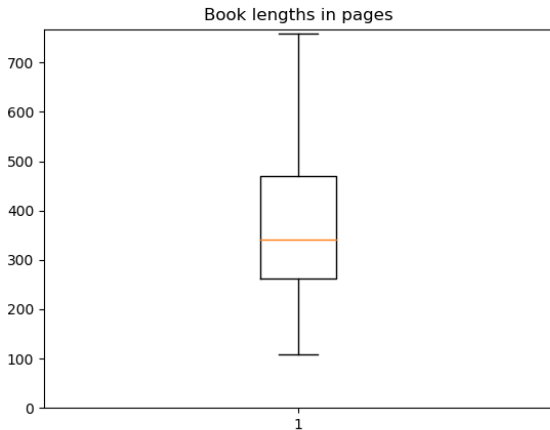
Show more



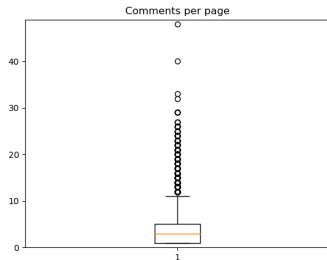
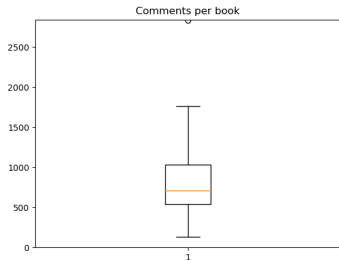
› Westedt hat in ihrer BA-Arbeit eine Stichprobe von Kapiteln hinsichtlich der Kommentare analysiert:

Kategorie	%
Stil	21.00
Wortwahl	20.73
Zeichensetzung	11.81
Grammatik	11.55
Referenzen	9.71
Satzbau	7.80
Rechtschreibung	7.30
<b>Inhalt</b>	<b>6.56</b>
Sonstiges	3.41

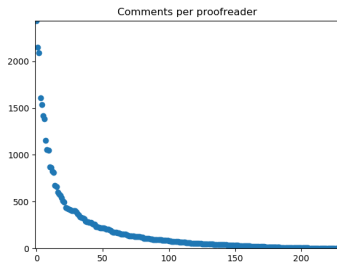
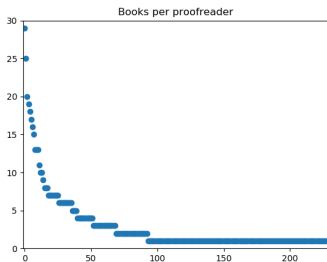
- › nach der qualitativen Analyse Westedts jetzt eine quantitative Analyse
- › 52 Bücher (2016–2018)
- › Paperhive-Kommentare wurden per API in eine Datenbank überführt
- › 19 004 pdf-Seiten
- › 43 370 Kommentare
- › Datenbasis verfügbar auf  
<https://doi.org/10.5281/zenodo.3063004>



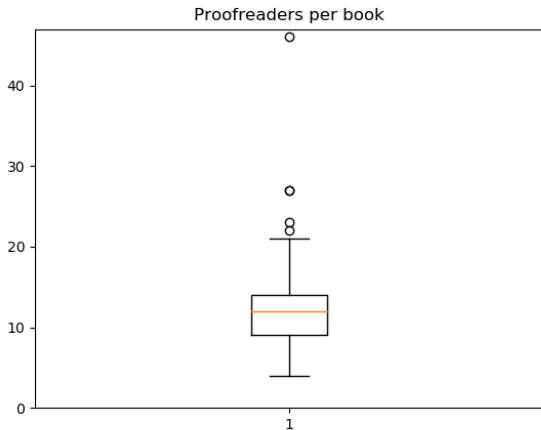
# Menge Kommentare



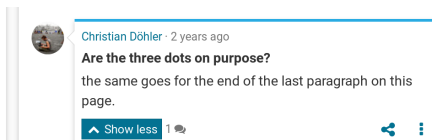
Die höchste Anzahl Kommentare auf einer Seite findet sich auf Seite 122 von *Theory and description in African Linguistics* (48 Kommentare).



228 verschiedene Accounts haben bisher am Community Proofreading teilgenommen.



course about the Jewish Nerwa texts, where I  
rstood it as a recommendation to further my  
this...

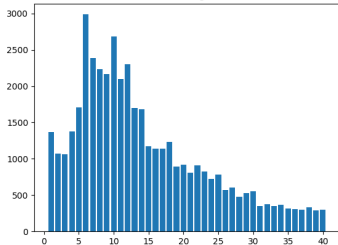


- › Kommentare auf PaperHive haben einen Titel (<40 Zeichen)
- › zusätzliche Ausführungen in Freitextfeld optional möglich.

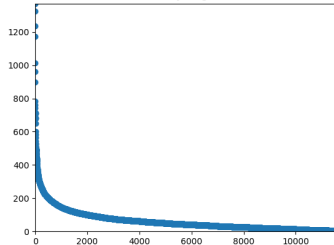


# Title length and body length

Title length



Body length

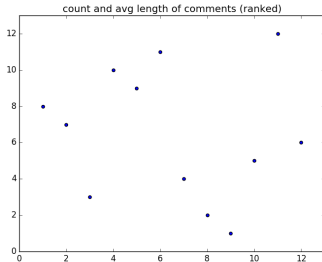


- › Community Proofreading liefert einen Datensatz an Kommentaren von beträchtlicher Größe und Güte
  - › Anhand des Datensatzes können verschiedene Hypothesen getestet werden
1. **Proofreader fallen in zwei Klassen: Typ 1 konzentriert sich auf kleine Details; Typ 2 sieht das Werk als Ganzes**
  2. **Kommentare werden gegen Ende des Kapitels kürzer (Ermüdung, Verweise auf vorher Gesagtes)**

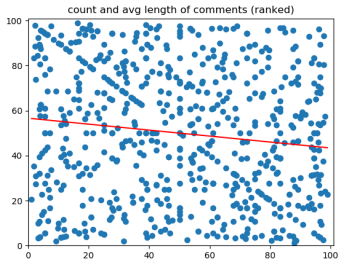
- › Typ 1: viele Kommentare, dafür kurz (“comma missing”)
- › Typ 2: wenige Kommentare, dafür ausführlicher

## › Für jedes Buch

- › ordne alle Teilnehmer nach Menge an Kommentaren
- › ordne alle Teilnehmer nach durchschnittlicher Kommentarlänge
- › trage beide Rangordnungen gegeneinander ab



- > 12 Teilnehmer
- > Rangfolge wird durch Punkte angegeben.
  - > e.g. #3 in Menge ist auch #3 in Durchschnittslänge, aber #1 entspricht #8
- > Daten eines Buches offensichtlich nicht ausreichend.



- › Ränge auf Perzentile normalisiert
- › Rote Linie gibt best fit an
- › in der Tat eine schwache negative Korrelation

## › Hypothese #1 bestätigt

- › Proofreader mit mehr Kommentaren haben kürzere Kommentare
- › Proofreader mit längeren Kommentaren haben weniger Kommentare

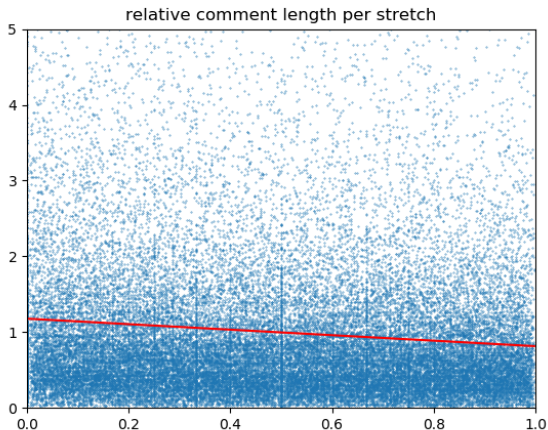
**Hypothese 2: Kommentare werden mit zunehmender Dauer der Korrektur kürzer, einerseits durch Ermattung, andererseits durch Verweise auf weiter oben Gesagtes.**



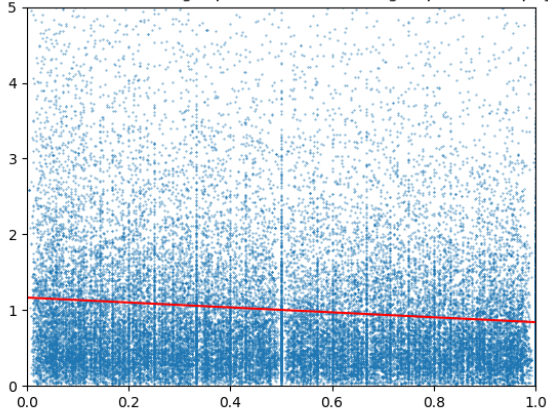
- › for every book
  - for every proofreader
    - for every comment
      - › ermittle relative Kommentarlänge in Prozent (zB 67% der durchschnittlichen Kommentarlänge)
      - › ermittle relative Kommentarposition (vorne, mitte, hinten) in Prozent
      - › trage die Ergebnisse gegeneinander ab
- › Ein Punkt (0.5, 5) bedeutet, dass der Kommentare in der Mitte des Kapitels 500% der Durchschnittslänge hatte.
- › Die relative Position kann anhand der Seiten oder anhand der gereihten Kommentare ermittelt werden

## Plot für Hypothese #2 nach gereihten Kommentaren

| Hypothesen  
| Proofreader fatigue



relative comment length per stretch according to position of page

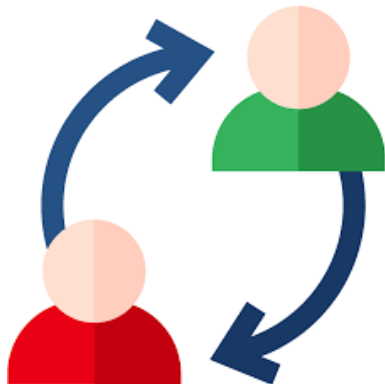


## › Die Hypothese ist bestätigt

- › je später ein Kommentare im Dokument auftaucht, desto kürzer ist er.
- › Der erste Kommentar hat im Schnitt 110% der Durchschnittslänge; der letzte Kommentar 90 %.
- › Effekt schwach, aber messbar.

- › Wesentliches Ziel ist methodologisch
- › Kommentare sind ein Nebenprodukt von Open Publishing
  - › im traditionellen Modell nicht zugänglich
- › Wenn Dokumente, Prozesse und Formate geöffnet werden, ergeben sich neue Forschungsfragestellungen, die in traditionellen Abläufen ausgeschlossen waren.
- › Eventuell interessant für die Leseforschung.

- › 908 Forschende mit Rolle  
“author” bei Language  
Science Press
- › 228 Forschende mit Rolle  
“proofreader”
- › 27 Forschende in der  
Schnittmenge
  - › 16 erst Autor, später  
Proofreader
  - › 11 erst Proofreader,  
später Autor
  - › Bewegung in beide  
Richtungen
  - › Bindung an den Verlag



- › Community proofreading ist ein neuer Weg zur Einbindung der Community
- › nur möglich bei OA
- › nur möglich bei Fachcommunities (nicht bei Uni-Verlagen)
- › Implementation für 50+ Bücher und 200+ Teilnehmer getestet
- › vergleichbar mit traditionellem Lektorat/Korrektorat
- › Nebenprodukte können für weitere Forschungsfragestellungen verwendet werden
- › Evolution in beide Richtungen Autor ↔ Proofreader
- › Forschende können sich je nach Erfahrung unterschiedlich beteiligen, um Manuskripte zu verbessern

- › Welche anderen Fragestellungen könnte man anhand dieser Daten beantworten?
- › Welche anderen wissenschaftlichen Disziplinen könnten interessiert sein?





## Gold proofreaders

› **Andreas Hölzl** ([view profile](#))



34/99

› **Jeroen van de Weijer** ([view profile](#))



31/99

› **Eitan Grossman** ([view profile](#))

29/99

› **Jean Nitzke**

25/98

› **Christian Döhler** ([view profile](#))

19/97

› **Martin Haspelmath** ([view profile](#))

19/97

› **Ahmet Bilal Özdemir** ([view profile](#))

19/97

› **Ikmi Nur Oktavianti** ([view profile](#))

15/96

› **Brett Reynolds** ([view profile](#))

14/96