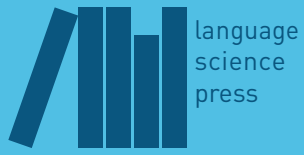
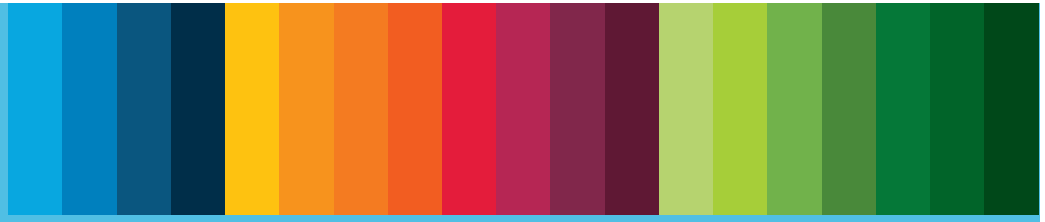




Community proofreading as a tool for community engagement

A quantitative analysis

Sebastian Nordhoff

May 30, 2019

FU Berlin

Open Publishing

- Open Access is mainly concerned with reading
- Open Publishing is concerned with making all aspects of publishing open
  - Open source platforms
  - Open bookkeeping
  - Open peer review
  - Community proofreading

Bibliodiversity

- one research can adopt different roles
  - author, reviewer, reader, ...
- junior researchers are more often readers
- senior researchers take on the other roles as well
- complex ecosystem
- community-based publishing tries to integrate researchers at all levels

Traditional proofreading

- outsourced work-for-hire
- for a fee
- one proofreader
- specialist in style and guidelines
- might have some training in linguistics
- normally no specialist knowledge of the particular subfield

Community proofreading

- crowdsourced to the community
- voluntary work
- many proofreaders
- very often specialists in the particular subfield
- intrinsic interest
- less acquaintance with style and guidelines

Language Science Press

- Open Access publisher in linguistics
- 100+ books since 2014
- 350 community proofreaders

Workflow

- proofreading queue with a new title every 2 weeks
- title is announced on Monday
- community members can volunteer and claim a chapter
- chapters are assigned on Wednesday
- 4 weeks time for proofreading
- proofreading is done on Paperhive

Westedt (2018)

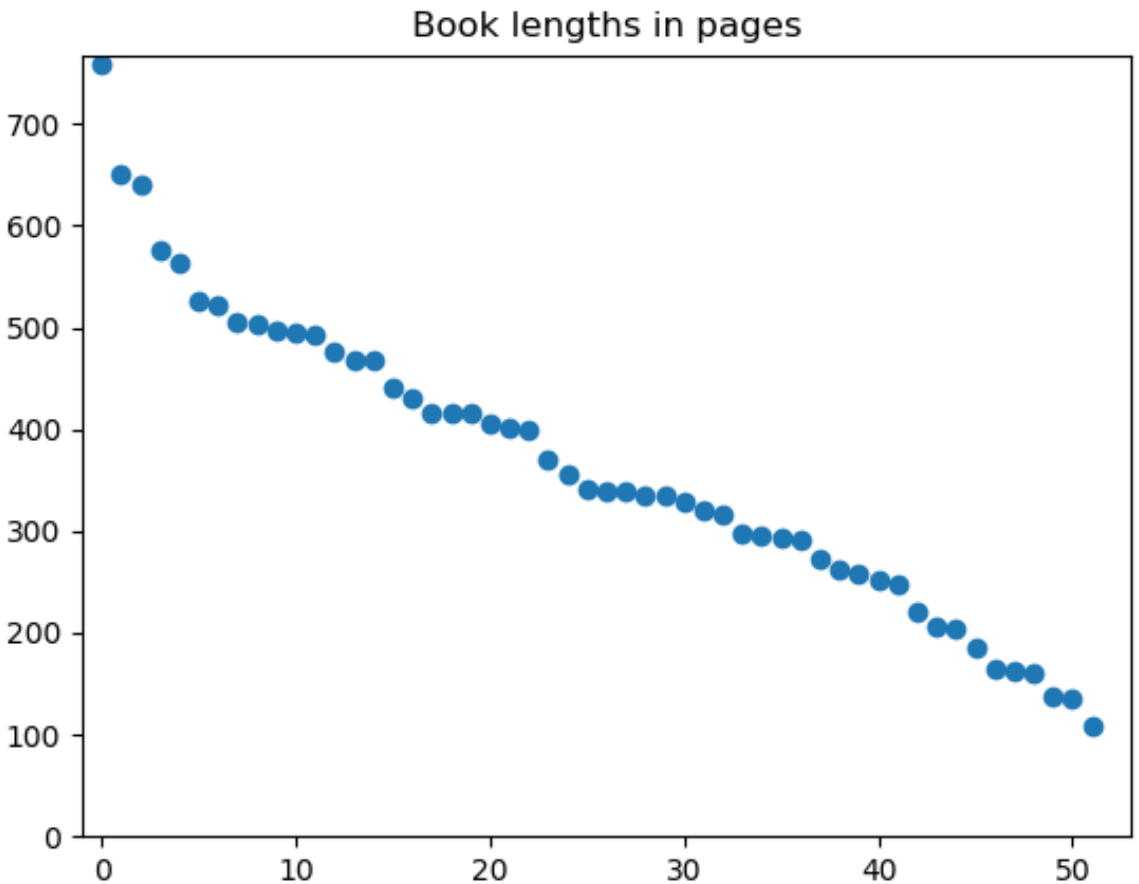
Westedt analysed a sample of comments on Paperhive for her BA thesis.

Category	Percentage
Spelling	7.30
Syntax	7.80
Lexical choice	20.73
Grammar	11.55
Punctuation	11.81
Style	21.00
Content	**6.56**
Miscellanea	3.41
References	9.71

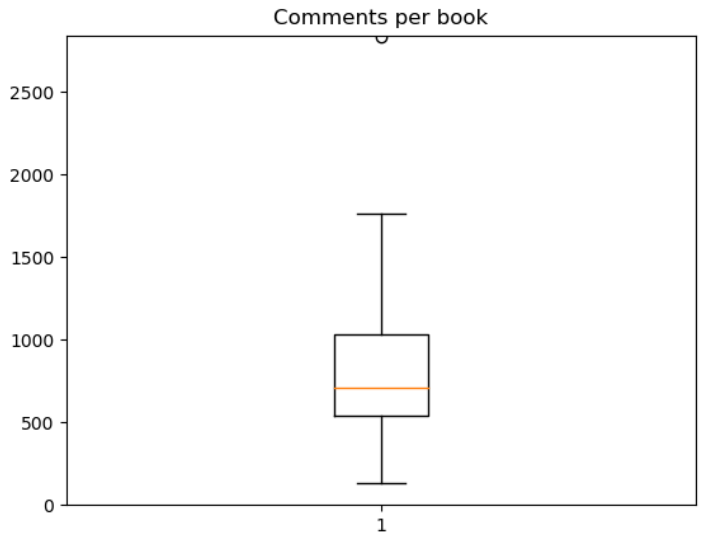
This study

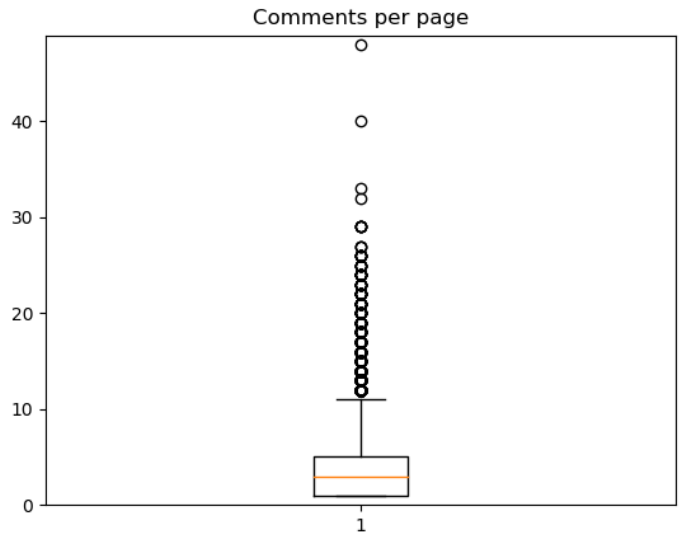
- 52 books from late 2016 to late 2018
- comments were harvested from Paperhive and put into a database
- 19 004 pages
- 43 370 comments

Book length


Book lengths in pages
700
600
500
400
300
200
100
0
0
10
20
30
40
50

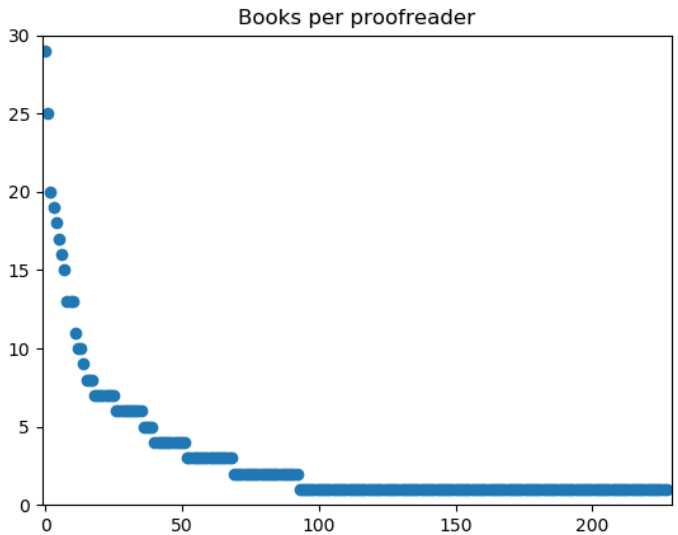
Comments

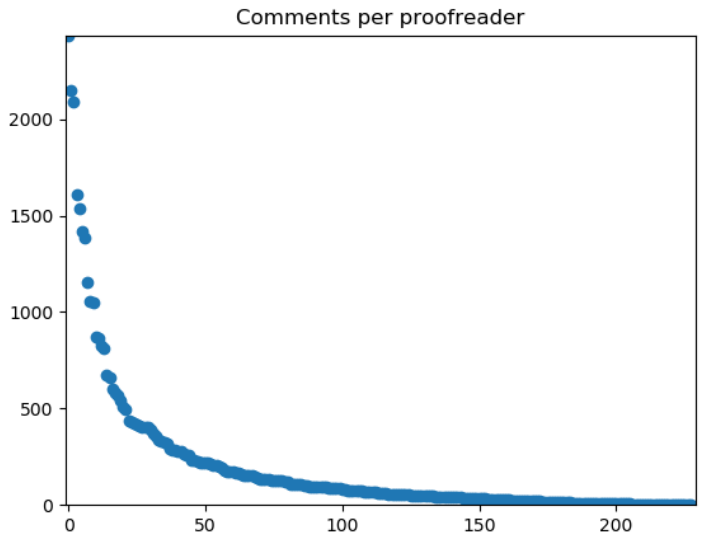




The highest number of comments on one page is found in Theory and description in African Linguistics on page 122 (48 comments).

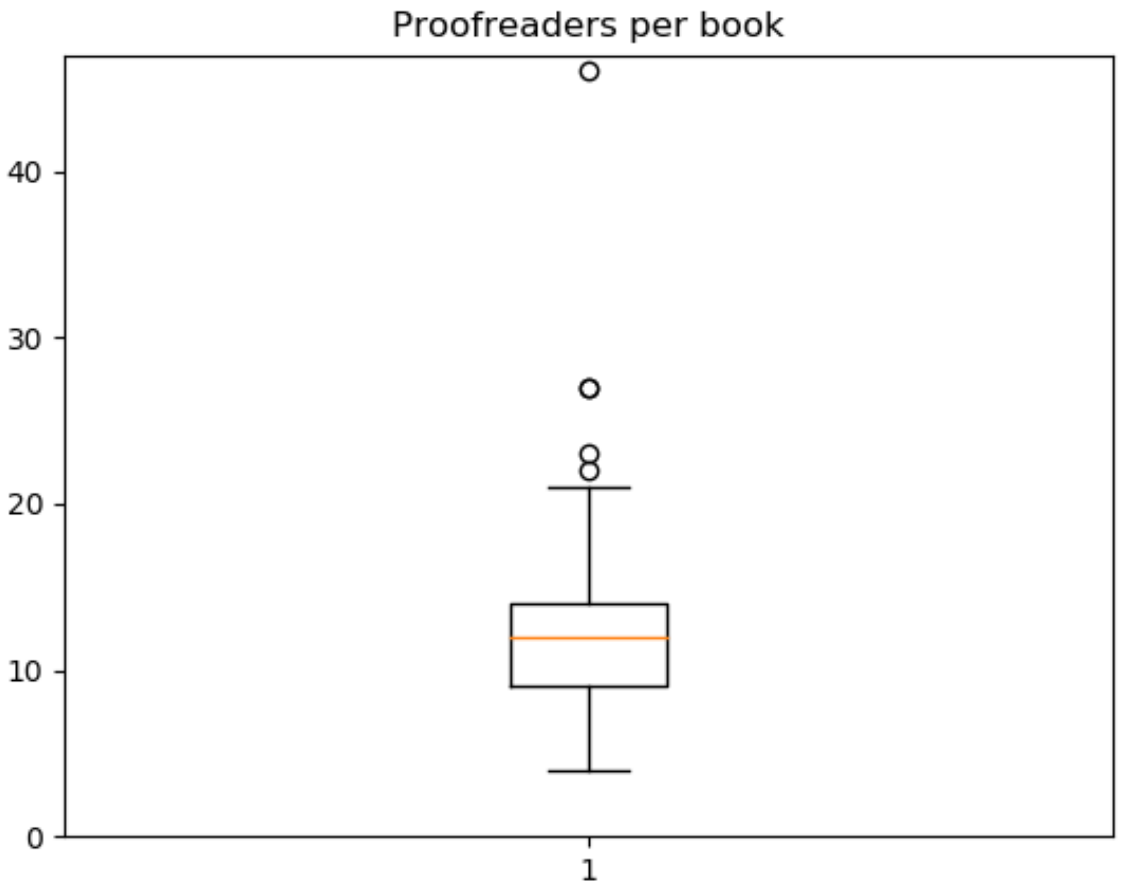
Productivity of proofreaders





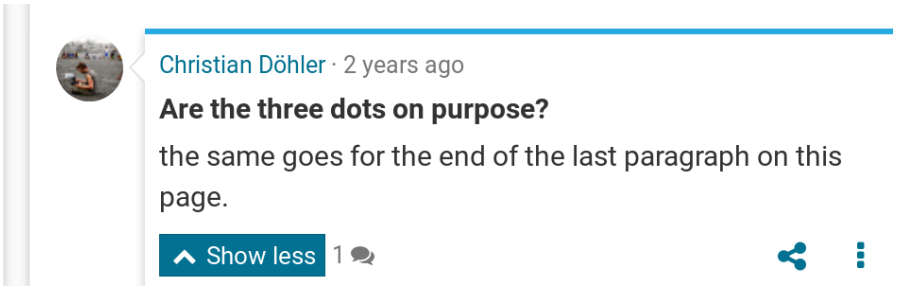
228 different accounts have participated in commenting.

Proofreaders per book



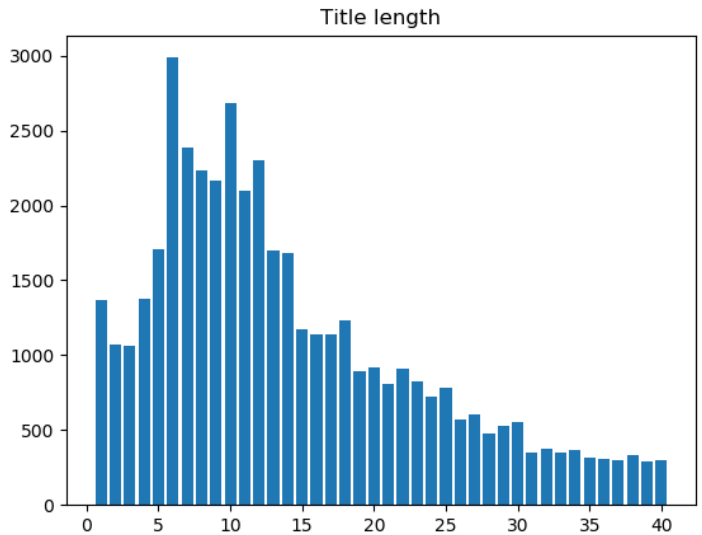
Text analysis

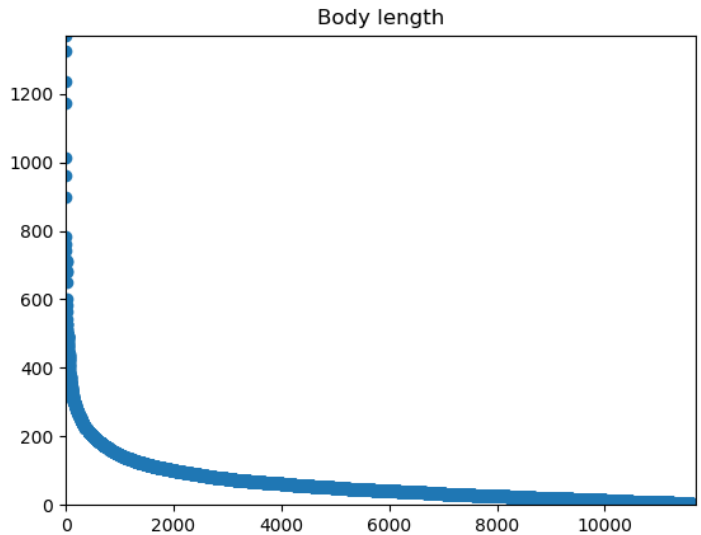
ourse about the Jewish Nerwa texts, where I rstood it as a recommendation to further my this...



- A PaperHive comment has a succinct title (<40 characters)
- optional body, with more elaborate information

Title length and body length





Hypotheses about proofreaders

1. **Proofreaders fall into two types. Type 1 will focus on small details; type 2 will focus on the big picture.**
2. **Proofreading comments will diminish as the proofreader moves along. Comments will become fewer due to fatigue, and average comment length will go down due to repetition of previous remarks as "see above".**

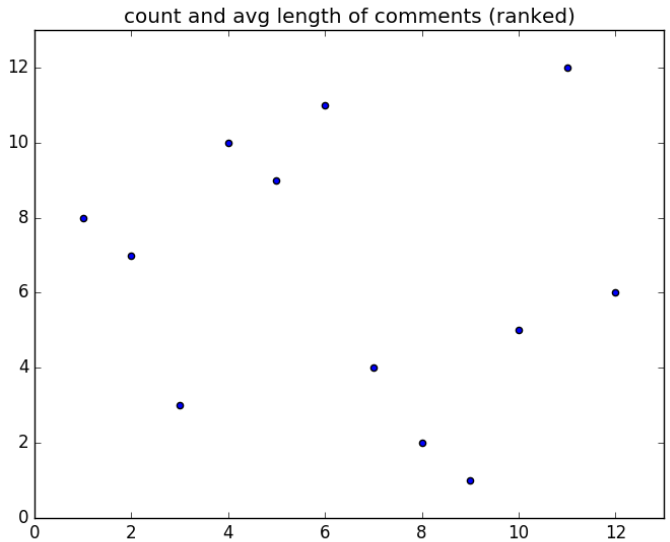
Hypothesis 1: proofreader types

- Type 1: many comments but short (“comma missing”)
- Type 2: few comments, but longer, in-depth

Computation

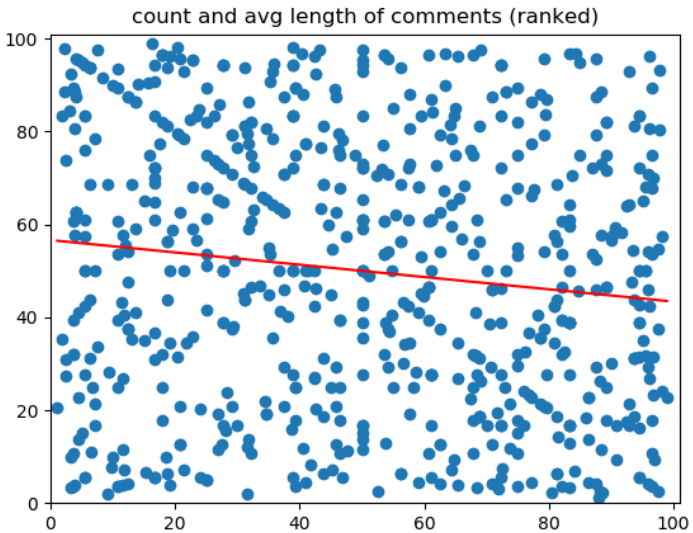
- For every book
  - rank all participating proofreaders by amount of comments
  - rank all participating proofreaders by average length of comments
  - plot the two against each other

Example of a plot for Hypothesis 1



- 12 proofreaders participated
- their respective ranks are given by the dots.
  - e.g. #3 in one rank is also #3 in the other, but #1 on one is #8 in the other
- data from one book insufficient

Combination of all books



- Ranks are normalized to centiles
- best fit given by red line
- indeed a weak negative correlation

Result hypothesis #1

- Hypothesis #1 is confirmed
  - proofreaders with more comments have shorter comments
  - proofreaders with longer comments comment less

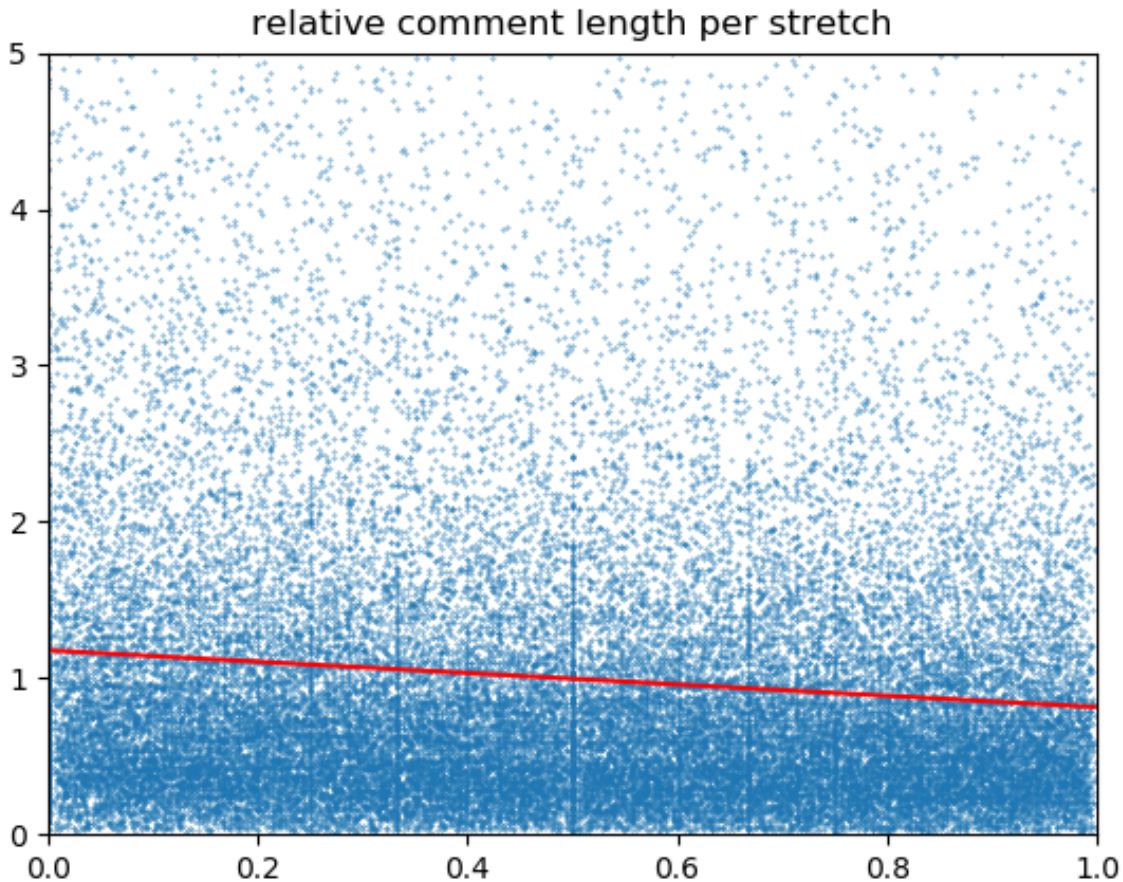
Hypothesis #2: reviewer fatigue

Hypothesis 2 : Proofreading comments will diminish as the proofreader moves along. Comments will become fewer due to fatigue, and average comment length will go down due to repetition of previous remarks as "see above".

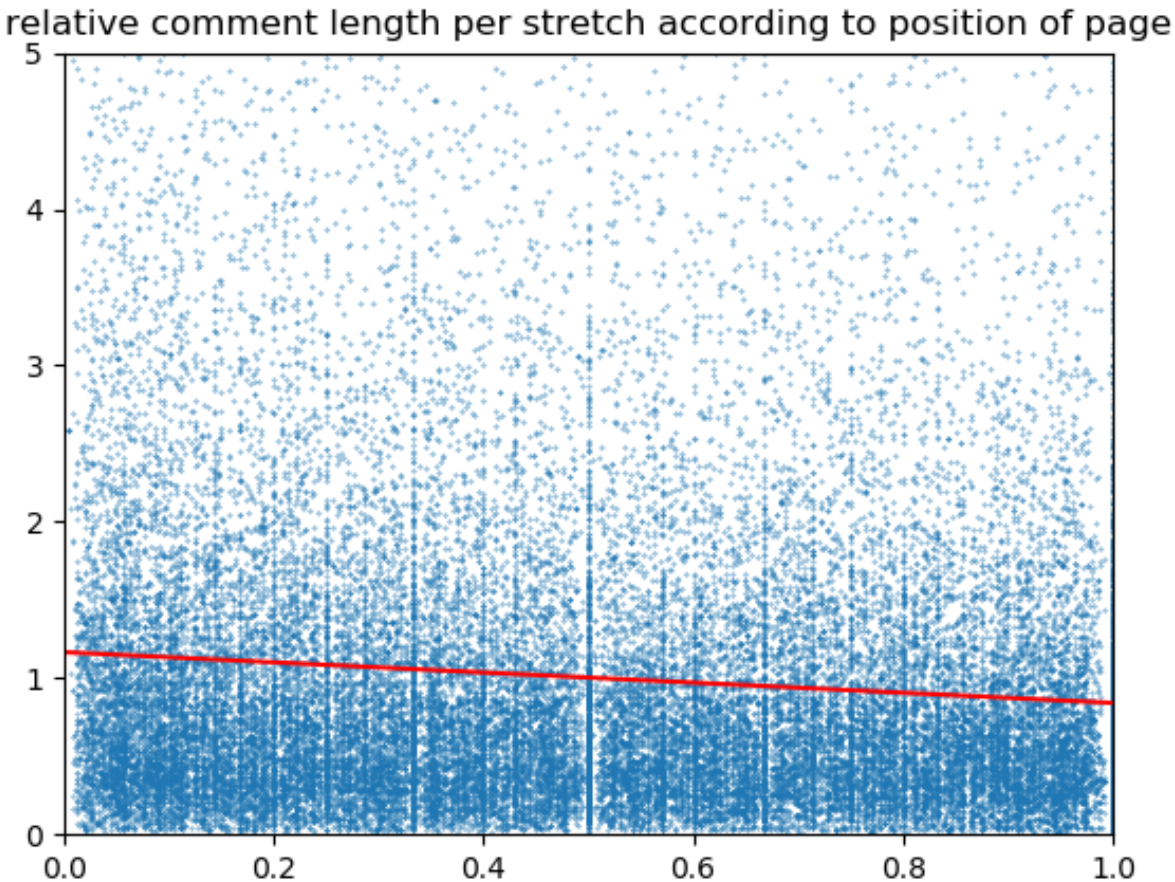
Computation for Hypothesis #2

- for every book for every proofreader for every comment
  - compute relative length (e.g. 0.67 of the average)
  - compute relative position (front, middle, back)
  - store the tuple (relative position, relative length)
  - A dot at (0.5, 5) means that there was a comment in the middle of the relevant stretch whose length was 5 times the average comment length.
- the relative position can be pegged to the linear order of comments, or to the pages

Plot for Hypothesis #2 based on linear order



Plot for Hypothesis #2 based on page position



Results for Hypothesis #2 “reviewer fatigue”

- Hypothesis is confirmed
  - the later in the document a comment is, the shorter it will be
  - the first comment will be about 110% of the average, while the last one will be 90% of the average.
  - effect not very strong, but discernible

Discussion

- Main aim: methodological
- Proofreading comments are a by-product of open publishing
  - In traditional publishing models, these data would not be available
- Once the documents, processes, and formats are opened up, novel research questions can emerge which would not have been possible under a closed setup.
- Implications for psychology of reading for instance.

Do researchers take on different roles?

- There are 908 people with the role “author” at LangSci Press
- There are 228 proofreaders
- 27 researchers have taken up both roles
  - 16 started as authors, and became proofreaders later
  - 11 started as proofreaders, and became authors later
  - Movement between the author pool and the proofreader pool in both directions.

Conclusions

-
-

Conclusions

- Community proofreading is a novel way of engaging the community
- only possible for Open Access publications
- workable implementation with 50+ books and 200+ researchers
- can compare to traditional proofreading
- by-product data can be used for novel research questions
  - proofreader typology
  - proofreader fatigue
- flow back and forth between the group of authors and the group of proofreaders
- healthy ecosystem
- researchers from different backgrounds at different stages of their career contribute their respective expertises to creating and improving manuscripts.

Thank you