Variation and overlap in the peer production of community rules

Sohyeon Hwang, Aaron Shaw
Northwestern University
Language editions as connected, self-governing community organizations

Goals
Technology infrastructure
Governance model

English
French
German
Japanese
Spanish
RQ1.) How do patterns of rule-making over time compare across language edition communities?

RQ2.) How do the sets of rules become more/less similar over time among language edition communities?
Descriptive study of nearly 20 years of rules and rule-making on five autonomous communities
Descriptive study of nearly 20 years of rules and rule-making on five autonomous communities

RESEARCH DESIGN

- Construct lists of rules per language community
- Collect revision histories and interlanguage links
- Descriptive analyses examining (1) comparisons across and (2) relationships between communities
RQ1.) How do patterns of rule-making over time compare across language edition communities?

RQ2.) How do the sets of rules become more/less similar over time among language edition communities?
KEY FINDINGS

Shared patterns of formalization

Figure 2: Trends in rule-related revisions (revisions on rule pages or their corresponding talk pages) by years since the first rule-related revision. 0 on the y-axis thus means the first year per language edition of rule-related revision activity.
Shared patterns of formalization

Decreasing size of revisions over time

Increasing time between revisions
Shared patterns of formalization

Several rule-making patterns observed in English Wikipedia replicate across non-English language editions.

Research and practical implications of the generalizability of findings + lessons from lots of work on English Wikipedia!
RQ1.) How do patterns of rule-making over time compare across language edition communities?

RQ2.) How do the sets of rules become more/less similar over time among language edition communities?
### Unique (0 ILL) vs shared rules (4 ILL)

<table>
<thead>
<tr>
<th></th>
<th>0 ILLs</th>
<th>4 ILLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule creation year</td>
<td>2007</td>
<td>2004</td>
</tr>
<tr>
<td>Rule page length</td>
<td>8,100</td>
<td>14,251</td>
</tr>
<tr>
<td># Wikilinks</td>
<td>449</td>
<td>31,806</td>
</tr>
<tr>
<td># Revisions</td>
<td>195</td>
<td>371</td>
</tr>
<tr>
<td># Unique contributors</td>
<td>53</td>
<td>234</td>
</tr>
</tbody>
</table>

34% 28%
Increasingly diverging sets of rules

Figure 6: The proportion of rules in a focal language edition in each year (X-axis) with the number of ILLs to the other language editions in the set (area shading).
Continued importance of shared rules

Figure 7: The proportion of rule-related revisions in a focal language edition in each year (X-axis) made to rules with the number of ILLs to other language editions in the set (area shading).
Divergence and convergence

Figure 6: The proportion of rules in a focal language edition in each year (X-axis) with the number of ILLs to the other language editions in the set (area shading).

Figure 7: The proportion of rule-related revisions in a focal language edition in each year (X-axis) made to rules with the number of ILLs to other language editions in the set (area shading).
Divergence and convergence in rules

Communities aligned in key ways develop substantial + sustained institutional variations in practice

Broad, widely-shared rules created early may help coordinate communities toward shared goals.
Shared patterns of formalization

Increasingly diverging sets of rules

Continued importance of shared rules
(which may be variably evolving)
Directions for future work

**How scale relates to rules and rule-making**
How do these patterns compare in mid-size and small language editions?

**Characterizing community-specific rules**
What kinds of rules are “unique”, and what makes them so?

**Varying enforcement of broad shared rules**
Do language editions use the same rule to different effects?
How can we address harmful effects?
Thank you!

Contact
sohyeonhwang@u.northwestern.edu
aaronshaw@northwestern.edu

Acknowledgements
• Support for this project came from the National Science Foundation (IIS-1910202, DGE-1842165).
• Special thanks to Brian Keegan, Casey Fiesler, Ágnes Horvát, Darren Gergle, and members of the Community Data Science Collective.
• A version of this work has been published and presented at ICWSM2022, DOI: 10.1609/icwsm.v16i1.19297