Do We All Talk Before We Type?: Understanding Collaboration in Wikipedia Language Editions

Taryn Bipat, David W. McDonald, Mark Zachry
Human Centered Design and Engineering
University of Washington,
tbipat, dwmc, zachry@uw.edu

ABSTRACT

The English language Wikipedia is notable for its large number of articles and for the intricate collaborative interactions that create and sustain it. However, 288 other active language editions of Wikipedia have also developed through the coordination of contributing editors. While collaboration in the English Wikipedia has been researched extensively, these other language editions remain understudied. Our study leverages an influential collaboration model based on behaviors in the English Wikipedia as a lens to consider collaborative activity in the Spanish and French language editions. Through an analysis of collaborative interactions across article talk pages, we demonstrate that talk pages, the locus of most collaboration on the English Wikipedia, are used differently in these different language editions. Our study raises broader questions about how results from studies of the English Wikipedia generalize to other language editions, demonstrates the need to account for variations in collaborative behaviors in all language editions of Wikipedia and presents evidence that collaborative practices on the English Wikipedia have changed overtime.

CCS CONCEPTS

• **Human-centered computing → Collaborative and social computing**; *Wikis*

KEYWORDS

Wikipedia; Collaboration; Multilingual; Talk Page; Content Analysis; Language Editions

OpenSym '18, August 22–24, 2018, Paris, France

ACM ISBN 978-1-4503-5936-8/18/08…$15.00
https://doi.org/10.1145/3233391.3233542

ACM Reference format:
T. Bipat, D. W. McDonald and M. Zachry.2018. Do We All Talk Before We Type?: Understanding Collaboration in Wikipedia Language Editions. In *Proceedings of the International Symposium on Open Collaboration, Paris, FR, August 2018 (OpenSym'18)*, 11 pages.

1 INTRODUCTION

Louis has been a very active Wikipedia editor on the French edition of Wikipedia for a long time and he is considering the possibility of becoming an administrator. He is a bit concerned if he has the credentials for promotion and so has begun to do research on the topic. He comes across a study by Burke and Kraut [3] that suggests a model for how editors in the English Wikipedia become promotable, and he is interested to see if it can be used as self-evaluation model. After comparing the model to his own experiences as an editor, he determines that he is ready for admin status, but he is disappointed when it never happens. He wonders why the model was not a more accurate predictor for his case.

Wikipedia, the fifth most popular website[1] on the Internet, is an online collaborative community built to give free access to encyclopedias in all languages. In total, 299 different language editions of Wikipedia exist; with 288 being active and over 20 of those have 100,000 plus articles. Three of the largest Wikipedia editions are the English, French and Spanish. English is the largest Wikipedia edition with over 5 million articles; French is the 6th largest with a little more than 1.9 million articles; and Spanish the 9th largest with 1.3 million articles[2]. Understanding the continuing development of this platform is literally of global interest. Much of the research on Wikipedia, however, has been conducted using the English edition. Consequently, our understanding of how editors aid in the growth of the world's premiere peer production community is based on practices in only one language edition, leaving open questions about norms in the other 287 active editions of the encyclopedia. Our study seeks to explore these disparities by qualitatively studying collaborative practice on three different language editions of Wikipedia: English (EN), French (FR) and Spanish (ES).

In order for editors to develop an in-depth and quality encyclopedia, they must collaborate. Wikipedia talk pages, also called discussion pages, are the most active site of collaboration

[1] https://alexa.com/topsites

[2] https://meta.wikipedia.org/wiki/List_of_Wikipedias

used by editors who wish to discuss edits that need to be made to articles. Each article has its own individual talk page and the size of these pages is dependent on the amount of collaboration occurring. Previous work has shown that prior to editing articles in EN, editors will first discuss the changes that need to be made on the talk page [28]. Talk pages have been so critical in the development of EN articles that archival processes have been created to help keep talk pages less convoluted and up to date. The WP:ARCHIVENOTDELETE[3] guideline page provides guidance and specific rationale for when and why to archive talk pages, preserving important rationale for future inspection.

While EN talk pages have been researched by many, we are additionally interested in whether collaboration in FR and ES occur similarly on talk pages to how it does on EN. Our exploratory analysis consisted of 48 articles, 16 similar articles across all three languages. We noticed that in EN there was an average of 8 talk pages (1 current and 7 archived), while in FR and ES, on average there was only 1 current talk page for the articles coded. We wanted to make sure the archival practices of ES and FR were not different from EN, so we analyzed the percentage of edits done on the talk page to the number of edits done on the article. EN had the highest proportion with 34.6%, while the ES and FR editions were much lower at 6% and 11.7%, respectively.

Prior research suggests that several forms of collaboration and coordination are visible mainly through talk pages in EN [16, 25, 28]. However, our exploratory analysis suggests these patterns may not be similar to those on FR and ES as there is much less collaboration occurring in those other language editions. Our study explores the similarity and differences in collaboration on talk pages across these three language editions. This study, however, is not designed as a direct comparison of three different language editions. Each of these platforms is different, reflecting its own culture and organizational structure, and these differences ultimately make them comparable in only a reductive sense.

Additionally, we do not want to add our own interpretations based on unknown assumptions about the cultural differences at play in each of these editions. We begin to foreground the different forms that collaboration takes in the cultures present in EN, ES and FR by using the same topics across different languages with similar organization structure and content. We then apply a collaboration model created from EN to allow us to better understand the inherent problem with generalizing Wikipedia research based in only a single language edition to potentially other collaborations on other editions.

2 RELATED WORK

Although most prior studies of editing practices in Wikipedia have focused on EN, researchers have recently started examining differences among the many language editions of Wikipedia [1,18]. These studies, however, have focused on content issues rather than underlying behaviors of editors in these different language editions of Wikipedia.

2.1 Behavioral Patterns on Wikipedia

In our motivating scenario, Louis, the avid editor on FR mentioned earlier, discovered that looking into research on what makes for a successful administrator candidate did not help him achieve promotion. Could this be because the research he consulted was based on the English Wikipedia and what holds true there differs from what is true in FR? Do studies of behavioral patterns on Wikipedia derived from behaviors in EN idiosyncratically represent what occurs in that language edition?

Many types of behavioral patterns in Wikipedia have been modeled in previous studies, but a common thread among these studies is their reliance on observable behaviors in EN. These EN-based studies include not only behavioral patterns related to administrator promotion processes [3], but also deletion processes [7] and content production processes [13]. Additional EN-based studies have proposed behavioral models of collaboration to satisfy technical and social requirements [12, 28]. Such models are further seen in studies of contributions on talk pages [13, 20, 25, 28], in the use of policies [14], in the use of Wikiprojects for coordination [19], and to recommend new talk page tools for editors [15, 25].

Collaboration models, primarily based on studies of EN, suggest recommendations to extend the system. For example, Lam et al. recommend that enhancements be based on how to influence the decision quality of the editors [15]. Kittur et al. suggest development of tools that can increase coordination and reduce conflict [11]. However, it is not evident to us that recommendations derived from behavioral observations in EN can be generalized across other language editions of Wikipedia. Given our interest in understanding collaboration, we leveraged an existing analytical coding scheme developed on EN [28] to begin to examine behaviors in ES and FR.

2.2 Multilingual Wikipedia

Research to better understand differences among Wikipedia language editions has increased. This turn to other languages is motivated by the need to understand Wikipedia's "neutral point of view" (NPOV) policy[4] across language editions. That policy asserts that all content must be written from a neutral point of view, "representing fairly, proportionately, and, as far as possible, without editorial bias, all of the significant views that have been published by reliable on a topic." Across language editions, NPOV helps ensure content bias arising from language differences is reduced, a potential bias that the community refers to as linguistic point of view (LPOV) [18]. Hecht & Gergle measure the multiplicity of topics in 25 different Wikipedia editions to prove the *global consensus hypothesis* false [9]. That is that the encyclopedic world knowledge is diverse across different languages and cultures, which in turn violates NPOV because the

[3]WP:ARCHIVENOTDELETE: https://en.wikipedia.org/wiki/Wikipedia:ARCHIVENOTDELETE

[4]WP:NPOV: https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view

same information can be represented different across language editions. Moreover, a study by Callahan and Herring demonstrates that articles about famous people exhibit differences in cultural and historical perspectives between Polish and English Wikipedia articles and also shows an English language content advantage both in the length of content and quality [5]. Another study focuses on measuring self-focus bias on Wikipedia across 15 different language versions [10]. This study casts further doubt on the *global consensus hypothesis* that encyclopedic knowledge is constant across languages; instead, highlighting that each language version offers unique information. Systems such as Omnipedia and Manypedia compare and analyze content from various language editions of Wikipedia [1,18]. These systems allow users to actively seek "*information exclusive to unfamiliar language editions and strategically compared how language editions defined concepts*" [1]. Manypedia enables LPOV and facilitates comparisons of content across the language editions, helping highlight content bias for Wikipedia users and editors. Research on multilingual Wikipedia content has gone as far as tying the language and content differences to cultural factors [18].

Most of this prior research shows that the English Wikipedia has a volume of content advantage, but every language edition has unique information [5,10]. This unique information is due to factors including linguistic issues, translation issues and cultural issues. For the purpose of this study we chose relatively similar articles in an attempt to mitigate differences across the multilingual talk pages resulting from content differences.

Few studies have looked at the behavior of contributing editors on Wikipedia across language editions. Nemoto and Gloor [22] compared conflict resolution in five different languages. Using a social network analysis lens, they noted that countries with lower Human Development Index such as Russia and Poland show less interest in editing and maintaining Wikipedia than more developed countries such as Denmark and Germany. Another study analyzes editing conflicts to demonstrate that Wikipedia is more than an encyclopedia but rather a social sphere with a variety of different interests and preferences [31]. These studies analyze editor behavior by interpreting article edit history, editor talk pages, and regular talk pages and then tie their findings back to the geographical region rather focusing on the collaborative and coordinative implications of the editors' work.

3 RESEARCH QUESTION

Calls for replication studies in the human-computer interaction research community have been prominent for the last decade. In 2011, for example, Wilson et al. [30] observed a bias in the community towards studies that report novel contributions. However, they point out that replication studies are necessary to confirm whether prior results generalize past specific samples and to make sure that assumptions hold from older research [30]. The call for more replication studies was repeated in 2014 [29].

To answer this call, our study extends multilingual Wikipedia research by attempting to replicate a study by Viégas et al. [28] that proposed a well-known collaboration model for EN Wikipedia. We extend that model to the collaborative practices in two other language editions of Wikipedia, FR and ES. The Viégas et al. study focuses on English Wikipedia talk pages and notes that most collaboration occurs in the talk page before becoming edits on the actual Wikipedia article. In our study, we use their coding scheme to examine collaboration across 16 articles from the three languages, analyzing 48 total articles.

Inherently, a replication of a study performed at different points in time, carries elements of a longitudinal study. While our research questions are not explicitly motivated by longitudinal issues, our attempt to replicate the Viégas et al. result provides the opportunity to raise questions about how EN Wikipedia collaborative practices have changed.

As we report here, we discovered that the ratio of talk page edits to article edits in EN versus FR and ES was so different that an additional investigation was warranted. If each language edition is unique, grows at different rates, and reflects its own cultural perspective, we believe that the collaboration model we see in [28] will not generalize across FR and ES. To investigate this, we conduct a content analysis of Wikipedia talk pages in EN, ES, and FR to determine how and what types of collaboration occur in different language editions of Wikipedia.

4 METHODS

4.1 Phase 1: Replicating Viégas et al. Coding Scheme

4.1.1 Generating Datasets. We built datasets of talk page contributions and applied the Viégas et al. coding scheme to it. We constructed our dataset by choosing articles that were similar in content and structure across all three languages. We selected articles across a range of topics to make sure there was a diverse dataset. We included articles that contained, (1) events that spanned countries where all of these languages were spoken, (2) articles on topics of universal importance and (3) articles that were also used in the prior study. We identified articles of global importance as those that had no unique ties to countries that spoke the languages analyzed. We followed this approach based on the assumption that influence by one language group on the topic can have impact on the collaboration behavior and content of article and so articles were chosen to avoid this conflict. Criteria 1 and 2 were selected to remove any bias towards the specific languages chosen in this study and criteria 3 to ensure generalizability and adhere to the prior methodology.

As shown in Table 1, every EN article had multiple talk pages (current and archived pages), while FR and ES typically had 1 current talk page except for the Religion, Earth, and Science articles. Mirroring the method in [28], we selected one current or archived talk page for each article to include in the dataset. To select the talk page, we chose the most recent talk page with a total length (by byte count) comparable to the average size of all talk

Article Title	EN	FR	ES
2014 FIFA World Cup (1)	6	3	1
2016 Summer Olympics (1)	3	1	1
Sun (2)	10	1	1
Religion (2)	12	2	4
United Kingdom (2)(3)	34	1	1
Tropical Cyclone (2)(3)	6	1	1
X (2)(3)	2	1	1
Hanged, Drawn and Quartered (2)(3)	9	1	1
Psychology(2)	3	1	1
Earth(2)	16	1	2
Science(2)	9	4	2
Carbon(2)	3	1	1
Water(2)	3	1	1
Photosynthesis(2)	3	1	1
Physics(2)	9	1	1
Mitosis(2)	2	2	1

Table 1: Number of current talk + archived talk pages in English, Spanish and French for the articles coded

pages for that particular article. The archival process for FR and ES are similar to EN, where talk pages are archived when they become lengthy. We selected a given article talk page on length to avoid situations where a recent archiving resulted in minimal talk page content.

We then gathered all individual posts from each talk page, using signatures and indentation levels to identify individual posts. In addition to the individual user posts on the talk page, we also included the template information boxes at the head of each talk

	EN	ES	FR
First Post	1/22/2003	10/29/2002	1/22/2003
Last Post	12/8/2017	12/18/2017	10/16/2017

Table 2: Dates of when the first and last post collected was created for each of the three languages.

Posting Dimension	Description
Requests/Suggestions for Editing Coordination	Postings that help users plan editing activity
Requests for Information	Postings where the user requests information that is related to the article topic without a clear intention to edit the article itself
References to Vandalism	Postings that refer to acts of vandalism on the article page
References to Wikipedia Guidelines and Policies	References were counted whenever users pointed out official Wikipedia guidelines either by name or by linking to policy pages
References to Internal Wikipedia Resources	Postings that link to other Talk pages or archives to illustrate/explain a user's comment
Off-Topic Remarks	Postings unrelated to the article.
Polls	Voting sessions organized by users to decide on controversial editing actions
Requests for Peer Review	Users hoping to elevate articles to "featured" status may solicit a peer review
Information Boxes	These are call-out boxes placed at the top of the Talk page indicating any special processes the article page may have gone through
Images	Image files posted on the Talk page.
Other	Postings that do not fit any of the above categories.

Table 3: Comments were classified into these 11 posting dimensions taken from Viégas et al. [28]

page. According to the Wikipedia talk page guidelines, archived talk pages include the information boxes in the original talk pages, but should not include the active talk page templates. For each analyzed talk page, we counted each information box as an individual post. In total, we collected 3407 talk page posts, 1924 posts in EN, 738 posts in ES and 745 in FR. Table 2 highlights the time periods when the posts collected were created.

Lastly, we translated the posts collected from the ES and FR talk pages to English. The first author's language proficiency in all three languages and the use of Google Translate assisted with this process.

4.1.2 Content Analysis. We collected exploratory data for each article and talk page, including counts of how many posts were in each talk page, how many talk pages existed, and byte counts. To conduct our content analysis, we used the coding scheme created by Viégas et al. [28] for 16 different articles in English, Spanish and French. As shown in Table 3, our coding scheme included 11 classifications as defined by Viégas et al. The codes used from Vièǵas et al. were partially mutually exclusive; however, because their study does not specify which codes were mutually exclusive, we used a mutually exclusive coding scheme to align our results as closely as possible with theirs. For posts that could have possibly been coded as multiple posting dimensions, the authors chose the code that covered the majority of the post content.

Three researchers met weekly to discuss coding to make sure the coding remained consistent with the original study.

Posting Dimension	Description
Interpersonal Conflict	Bickering with other member, directed at another user. For example, a personal attack or insult.
FYI	Posting is a statement or an announcement as a fact--contains no requests or suggestions
Opinion	Opinion user formed about anything with no requests or suggestions. May include agreement / disagreement among users
External Sources	Cites sources outside of Wikipedia.
Acknowledgements	Courtesies to acknowledge another user's post
Action	Statement that an action was taken or not taken
Other 2	Does not fit into any other code; rarely used

Table 4: The posts classified as Other in part 1 were categorized into these 7 posting dimensions

Additionally, we had multiple members of the research team re-code select sections of the data.

4.2 Phase 2: Unpacking the "Other"

Prior work has shown that not all discussions on talk pages is associated with article editing activity [6]. We coded such contributions as *other*. Subsequently, some of the *other* content also does not fit into the posting dimensions in Viégas, et al. Similar to [19, 25] we thus added codes to the scheme to enhance the original coding scheme. After testing the robustness of the Viégas et al. scheme in phase 1, we determined that a deeper analysis was required to understand what was going on in these talk page contributions. We thus recoded the data in our *other* section and created our own coding scheme as shown in Table 4.

Motivation to develop our coding scheme was based on deeper analysis of the posts classified as *other* in the first coding phase. The first author developed a coding scheme based on common themes across these posts and prior research. The codes may reflect prior research, for example, FYI [19] and External Sources [25]. However, these codes were picked because they are grounded in the dataset and were not drawn directly from other work. The schematic was later refined through additional coding by another research team member and further discussion by the entire research team.

5 FINDINGS

An effective collaborative pattern on EN, according to [28], is for an editor to first explain on the talk page which changes are going to be made to an article, wait a short time, then edit the article to make the specified changes, and to finally state on the talk page that the proposed changes have been made. This pattern would result in a 1:2 article to talk page edit ratio. Based on our data (show in Table 5), the 1:2 edit ratio as the normative collaborative practice in the

	EN
Edits on Target Article	7760
Edits on Talk Pages	2684.7
Size of Target Article (Bytes)	118.5 KB

	ES
Edits on Target Article	2910.6
Edits on Talk Pages	174.4
Size of Target Article (Bytes)	91.8 KB

	FR
Edits on Target Article	1748
Edits on Talk Pages	204.4
Size of Target Article (Bytes)	93.4 KB

Table 5: Each table consists of the overall edits for target article and related talk pages and target article sizes (in bytes) for EN, ES and FR.

EN no longer holds and, based on our data, was probably never true for FR and ES.

The findings for phase 1 (shown in Table 6) suggest that since the 2007 paper by Viégas and al. collaboration patterns in EN have changed. In EN, the category *Requests/Suggestions for Editing Coordination* is now the second largest category being surpassed by *Other*. However, among the non-*Other* categories, *Requests/Suggestions for Editing Coordination* is still the largest category in EN. Talk pages still appear to be the place where editors are making attempts to collaborate.

However, based on the number of edits to a talk page; talk page edits are to a talk page; talk page edits are no longer representative of the 1:2 ratio seen in the original study. Editors are spending more time, more text, discussing the topic of the article rather than explicitly discussing the edits they are making as illustrated by this editor's comment:

> "The South Florida Water Management District is a regional governmental agency responsible for water quality, flood control, water supply and environmental restoration in 16 counties, from Orlando to the Florida Keys. It is the oldest and largest of the state's five water management districts."

This editor does not mention any intention to edit the article but rather shares factual information about the Florida government's role in water management. Akin to the experiences of the editor of the comment above, we observed a shift in EN from discussing how to improve the quality of the article through editing

	Viégas et al. Average [28]	EN Average	ES Average	FR Average
Requests/Suggestions for Editing Coordination	58.8%	35.1% (676)	36.6% (270)	37.2% (277)
Requests for Information	10.2%	9.6% (185)	11.4% (84)	8.3% (62)
References to Vandalism	8.5%	0.9% (17)	0.7% (5)	0.6% (4)
References to Wikipedia Guidelines and Policies	7.9%	2.3% (44)	1.0% (7)	2.8% (21)
References to Internal Wikipedia Resources	5.4%	2.7% (52)	1.1% (8)	1.8% (13)
Off-Topic Resources	3.5%	0.8% (15)	0.5% (4)	0.3% (2)
Polls	0.4%	0.2% (4)	0.0% (0)	0.0% (0)
Requests for Peer Review	0.3%	0.9% (15)	0.6% (5)	1.6% (12)
Information Boxes	1.6%	6.2% (119)	15.8% (117)	11.3% (84)
Images	0.2%	1.2% (23)	0.0% (0)	0.1% (1)
Other[6]	3.2%	40.2% (774)	32.3% (238)	36.0% (269)

Table 6: Average from Viègas et al. [28] and overall averages of posting dimensions across language editions. The total amount of talk page posts per posting dimension is included in the parentheses.

coordination to debating the subject of an article. *Requests for information* is the next largest posting category in EN; it has not changed notably since the prior study with a decrease of 0.6%. In Viégas et al. *References to Vandalism* was the third largest posting category but in current practice, such postings have seen a substantial reduction from 8.5% to 0.9%, making it one of the least prominent contribution types on an EN talk page. This reduction may be attributed—at least in part—to the proliferation of bots in Wikipedia, which now address malicious contributions more extensively than they did at the time of the original study. The shift to discussing the content of articles rather than specific edits to the article may also explain the decrease in *References to Wikipedia Guidelines and Policies* postings. The following quote demonstrates how more discussion revolves around implicit referencing of the guidelines/policies:

> "That was in 2012. This is now 2014, when consensus and enforcement may be different. An event such as a sporting event that lasts several weeks is not really "in progress" when it is an off-day, or the local time is in the wee hours of the morning, when it is highly unlikely that there may be sudden rapid changes during those hours. If we did that, thousands of articles would have current events templates constantly on 24/7 for several weeks without informational consequence."

Rather than explicitly naming the ***Wikipedia Guidelines Template: Current Sport***[5], this editor offers a discourse about the guidelines. In direct contrast, our findings show an increase in the *Information Box* category.

Part of this increase can be accounted for by the common practice of creating a template to explain formal guidelines or some state of the talk or article page. These templates are placed on the page to simplify explanations (as seen in Figure 1) rather than users explicitly pointing to such guidelines in their individual posts. While participants were less likely to explicitly state guidelines and policies, there has been an increase in the use of templates on talk pages.

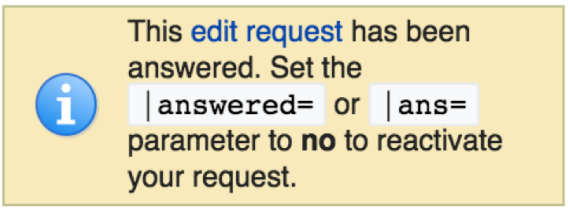

Figure 1: A template information box that contains talk page guidelines

[5] https://en.wikipedia.org/wiki/Template:Current_sport
[6] [28] did not report the totals in the *Other* category. We reached out to clarify what seemed like a discrepancy in the total percentages for their categories and they confirmed that the missing values should be attributed to *Other*.

Our findings demonstrate an increase in users discussing the subject matter of an article rather than the edits being made to the article. The decline in *References to Internal Wikipedia Resources* and *Off-Topic Resources* categories further support this trend. The remaining non-*Other* posting dimensions--*Requests for Peer Review, Images* and *Polls*--account for a total of 2.3% of the postings and differed by 1% or less from the results of the prior study. Overall, our findings show that since the Viégas et al. study was conducted, the use of talk pages in EN has changed. Talk pages[7] first emerged as a space for discussion about editing an article. Contrary to this original purpose, talk pages seem to be transitioning to accommodate Wikipedia users who are interested in discussing a particular interest with like-minded users. The largest posting dimension is the *Other* category--the category with the largest increase since the original study. This increase might be attributed to the shift in editor behavior on Wikipedia and gave us reason for investigating the editor behaviors represented in the *Other* category.

Overall, comparing the size (by bytes) of the talk pages that we coded and the numbers of edits, EN is much larger than FR and ES. Analyzing the posting dimensions, each language exhibits its own collaboration patterns. The differences in these collaborative behaviors may seem subtle but are observable differences, especially in some dimensions (see Table 6). Overall, *Requests/Suggestions for Editing Coordination*, *Requests for Information*, *References to Wikipedia Guidelines and Policies* and *Information Boxes* were the four posting categories with the biggest impact on the foreign language collaboration models. In comparison to the other two languages, FR had the most posts classified as *Requests/Suggestions for Editing Coordination and References to Wikipedia Guidelines and Policies*. The greater frequency of posts in these two categories may suggest that ES talk pages follow the rules originally created for talk pages and that its editors are still focused on the editing work of Wikipedia. In the following quote, an editor of the FR focused more on suggested what edits needs to be made to the article:

> "Hello, in this article, it is stated that the atoms forming linkages 6 (instead of the usual 4) was observed for the first time. If anyone is interested, it can be added to the article."

Similar to this quote, most of the contributions in the FR page went into elaborate detail about why the requests and suggestions for editing coordination were necessary to the curation of the article. We also noticed differences in the frequency of information boxes in talk pages across the three Wikipedia editions. These differences may be attributed to stylized distinctions that we saw in the language editions, in particular FR. As shown in Figure 2, we saw a different organization scheme across the talk pages in FR. While editors in EN and ES use indentation to demarcate between individual posts, FR has implemented technical enhancements that manifests as additional boxes outlining the threads.

Across all the languages, one of the most common posting type was *Other* (see Table 7). To better understand this observation,

Figure 2: The FR talk page for the 2016 Summer Olympic article includes special boxes to organize the threaded discussion

we implemented phase 2 of our methods. The topmost portion of the *Other* posting dimension was *FYI*, which for all three languages consisted of around 40% or more. This frequency seems to indicate that talk pages in all three languages contained a greater amount of discussion around the topic of the target article. The second largest category for ES and FR was *Opinion.* Both ES and FR had 20% or more of their *Other* postings in this category. This frequency of opinion posts in those two editions suggests that in both languages editors show more ownership over their posts. Unlike ES and FR, EN users had the lowest number of posts classified as *Opinion.* This low frequency of opinion posts and the fact that EN users had the highest number of *External Sources* related posts may imply that EN editors are less invested in their content.

The second largest category for the EN was *External Sources* which was also substantial for ES. This finding may demonstrate that ES sticks to traditional talk page policies and using external sources satisfies Wikipedia's verifiability policy[8] that anything that can be challenged needs to be cited. Additionally, ES has a large number of *Acknowledgements* posting which abides by Wikipedia's pillar that editors should treat each other with respect and civility. However, the large number of *External Sources* in EN was surprising because our results from phase 1 suggest an overall decrease in the frequency of policy citations in EN talk page posts. We suspect that this increase is connected to the *Interpersonal Conflict* category. As noted by Kriplean et al., citations were used to defend an editor's perspective on delimiting the scope of the article [14]. Furthermore, research has shown that external sources have been used in talk pages to express disagreement with the beliefs of the other editors [2]. In the following quote, external sources were used as a way for the editor in EN to support their argument when interpersonal conflict arose:

> "This article is about being hanged, drawn and quartered in general, so why is it important which law they were found guilty under? And yes, they were both found guilty of high treason. Felim O'Neill [1] [2], Robert Emmet [3], Two brothers sentenced in Ireland to be hanged drawn and quartered for high treason [4]."

[7]WP:TALK: https://en.wikipedia.org/wiki/Wikipedia:Talk_page_guidelines

[8]WP:VERIFY: https://en.wikipedia.org/wiki/Wikipedia:Verifiability

Other Posting Dimensions	EN Average	ES Average	FR Average
Interpersonal Conflict	8.5% (66)	3.7% (9)	4.6% (12)
FYI	43.9% (339)	39.3% (93)	47.0% (127)
Action	5.8% (44)	0.5% (1)	4.8% (13)
Acknowledgements	3.5% (27)	15.6% (37)	8.6% (23)
Opinion	11.5% (89)	19.9% (48)	30.6% (82)
External Sources	25.5% (197)	19.2% (46)	0.7% (2)
Other 2	1.5% (12)	1.8% (4)	3.8% (10)

Table 7: Overall percentages of each new category calculated out of the original *Other* category. The total amount of talk page posts per posting dimension is included in the parentheses.

This talk page user linked to outside articles to prove a point in a debate with other editors. The last two posting dimensions: *Action* and *Other 2* represented the smallest portion of the posting dimensions. The differences between these codes do, however, further support our contention that each of the languages have their own collaboration norms.

By unpacking the *Other* category, our findings further illustrate how each language's collaborative practices vary (see Table 7). The percentage differences for some posting dimensions are very small, however, and instead show some similarity between language editions. After phase 1, it was noticeable that posting dimensions with small numbers of posts were similar. *References to Vandalism, Off-topic Resources, Polls, Requests for Peer Review and Images* all had 1% or less difference in posts across the three language editions. The collaborative practices that show similarity across languages were the least prevalent in these three language editions. For example, *Requests for Vandalism* was similarly low for all languages. We suspect that this may be because each of the languages have their own version of a vandalism bot[9]. Additionally, we suspect that practices that were present in EN in 2007 such as polling or peer reviews are not as applicable to present day talk pages across the editions.

While we have demonstrated that some dimensions are strikingly similar, the overall collaborative practices in EN, ES and FR are not the same. This is further demonstrated in phase 2, the breakdown of the *Other* posts shows that while some percentage differences are small, the differences across the talk page behaviors are different.

6 DISCUSSION

Our study demonstrates that the collaborative practices across language editions of Wikipedia, specifically EN, ES and FR, are varied. Additionally, our findings draw attention to a temporal disparity between the findings of Viègas et al. and our data. Our initial research questions were not motivated by longitudinal differences; however, we highlight some implications that these findings may have for the future of research on collaborative platforms.

6.1 Evolving Behavioral Patterns on Wikipedia

Our study demonstrates that the collaboration model presented in Viégas et al. no longer accurately represents practices in EN. Most notably, practices in the *Requests and Suggestion for Editing contribution*, *Referencing of Guidelines and Policies* and *Referencing of Vandalism* categories have changed. In the last 11 years, Wikipedia has evolved in size, the nature and scope of its policies, guidelines, interface and other ways. These changes have led to new forms of collaboration and article development.

At the start of Wikipedia in 2001, there was an exponential increase in content and editors [8]. However, during March 2007, there was a turning point as the rate of page and editor growth of the platform began to slow down [26]. Moreover, with the rise and decline of the platform, the overall infrastructure has continued to evolve with the development of new policies and mechanisms for editing [4]. The decline in editing work has been partially attributed to the resistance of edits over time [26], the development of the platform's bureaucracy [4,26] and a decrease in active editors [8].

Halfaker et al. demonstrate how the development of platform bureaucracy to manage the overall increase of growth has, instead, led to a downturn in participation. The quality management mechanisms and algorithmic tools used to detect damage are experienced as being too restrictive for newcomers and, in turn, have led to a decrease in the number of active editors [8]. In addition, as online collaborative platforms continue to unfold, editor behavior and norms of the platform become more fortified over time [8, 27]. TeBlunthis et al. replicated the study proposed by Halfaker et al. in Wikia wikis to confirm that norms restrict the ideas of newcomers. Moreover, as more mature editors have authority over the system management, it becomes difficult for these norms to change [24, 27].

Holistically, the evolution of the bureaucratic nature of Wikipedia has influenced the motivation of editors and slowed the

[9] https://en.wikipedia.org/wiki/Wikipedia:Bots

production of content. As with the Wikia wikis, other Wikipedia platforms have the potential to be impacted by these bureaucratic developments. For example, it has been shown that Wikiprojects have changed over time [19]. Morgan et al. use a mixed-methods approach to demonstrate how there has been a behavioral shift from content creation to content curation that impacts the maintenance of quality content on the platform [19].

Collaborative practices that were prevalent in 2007 have changed. At that time, practices emphasized the use of talk pages to develop discussions about editing an article. In contrast, our 2017 findings show that since then these pages have transitioned to spaces that users with particular topical interests can engage in discussions with like-minded users. The bureaucratic development of Wikipedia that has offered stylistic differences, the propagation of bots, and increased use of new system features (e.g., extensions and protected pages) are all possible reasons why the Viégas et al. model no longer accurately accounts for editor practices in EN. Furthermore, Müller-Birn et al. note that social behaviors have been formalized into algorithmic processes such as bots [21]. These changes have led to a transition in the talk page space and open the door for the development and design of new collaboration models. Continued focus on forms of collaboration and coordination in Wikipedia is thus needed to increase the quality and quantity of the articles, but also to improve the user experience of editors. Consistent with Nieder and van Dijick [23], the amount of work done by system advancements such as bots across most language editions has had a great impact on the nature of the platform. Consequently, as researchers we need to further explore these developments to better understand how Wikipedia fulfills the social and technical needs of its users but also how the platform bridges linguistic and cultural gaps. Additionally, the insights developed from leveraging a collaborative model in EN from 2007 underscore the need for replication of prior studies of user behavior in peer production communities that are constantly changing. With the development of new underlying technologies on Wikipedia and changes in the adoption of these technologies, we speculate that many of the assumptions made in past research do not necessarily hold true as these editions evolve.

6.2 Generalizing across Wikipedia Language Editions

Across all three-language editions in this study, collaborative practices are unique. Ultimately across languages, differences in how much collaboration is occurring are evident. While it is possible to adapt models from prior Wikipedia research to other languages, as we attempted in this study, our results demonstrate that findings based on a study of one language edition cannot automatically be extended to other language editions. Research conducted on collaborative systems such as Wikipedia should not apply knowledge from an Anglo-centric (or another other language) model to other language editions. Previous work [13] has shown that collaboration is important to the creation of high quality articles. By understanding the different collaboration models across language editions, we can keep consistency in the quality of articles across all the editions of Wikipedia. Global system like Wikipedia must be understood as a sum of all of its parts--with the language editions being sub-sections that are unique. Users are a core driving factor in creating content so it is important that researchers treat Wikipedia as a sum of all of its parts rather than just focusing on EN, in order to better understand how to handle the fluidity of user experience across the various language versions.

Moreover, our findings suggest that the EN platform has its own set of characteristics that do not directly influence user behaviors in other language editions. In response, we question whether EN needs to be included as the comparator in multilingual research. Currently, most of the multilingual work rely on EN because of its expansive scope and the quality of its articles, but we call for future work to dig deeper into the other 288 active language editions of Wikipedia. Furthermore, we recommend additional qualitative work that analyzes more than 2 language editions. By focusing their research on only 1 or 2 language editions of WP, investigators can perpetuate a negative bias towards multilingual research in the CHI community. Currently, most of the multi-edition research with more than 2 editions are quantitative; however, we contend these studies do not do justice to how behaviors actually happen on these platforms. Quantitative, multilingual research can illustrate the organization and frequency of behavior but the collaborative patterns in different language editions are not numerically comparable. The collections of edit counts or edit sessions for a particular period of time do not represent the same behavioral patterns in different languages. Similarly, as shown in this work, the passage of time influences when and how behavioral changes occur and quantitative research cannot strongly reflect the time progression of user behavior in the editions. Additionally, by including the breakdown components of the behavioral organization, we can propose suggestions for why these differences occur.

7 LIMITATIONS

Our study has three main limitations. Our data shows very small percent differences between some of the coding dimensions. Additionally, we are a different group of researchers and can likely impart our own interpretations on the coding scheme. Both of these limitations fall under the nature of qualitative research, where one interpretation by a group of researchers may not be similar to another. The authors of this paper tried to reduce such bias by adhering closely to the method described in the Viègas et al. paper.

8 CONCLUSION

Our study examines whether collaboration practices differ across different language editions of Wikipedia, leveraging a prior, well-known model to do so. Our methods allowed us to assign an older quantitative coding scheme to 48 talk pages from the English, Spanish and French language editions of Wikipedia, to offer two contributions. We first have a better understanding of the impact age can have on collaboration models and secondly that behavior on language editions of Wikipedia are unique and thus we must be more aware of how we adapt findings from research on online collaborative systems from one language to another. Subsequent researchers may extend this work to apply social linguistic theories

to delve deeper into individual characteristics such as verbosity and language structure of each model and crowdsourcing habits of different language-speaking cultures to be able to link the linguistic findings with cultural knowledge. Additionally, future work should include further investigating how Wikipedia evolves over time and deepening our knowledge about how linguistic differences influence collaboration and equality in online communities.

ACKNOWLEDGEMENTS

We would like to thank Zena Worku and Jonathan Morgan for their input during this study. We would also like to acknowledge support from National Science Foundation (NSF) grant, IIS-1162114.

REFERENCES

[1] Patti Bao, Brent Hecht, Samuel Carton, Mahmood Quaderi, Michael Horn, and Darren Gergle. 2012. Omnipedia: Bridging the Wikipedia language gap. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12). ACM, New York, NY, USA, 1075-1084. DOI: http://dx.doi.org/10.1145/2207676.2208553

[2] Emily M. Bender, Jonathan T. Morgan, Meghan Oxley, Mark Zachry, Brian Hutchinson, Alex Marin, Bin Zhang, and Mari Ostendorf. 2011. Annotating social acts: authority claims and alignment moves in Wikipedia talk pages. In Proceedings of the Workshop on Languages in Social Media (LSM '11). Association for Computational Linguistics, Stroudsburg, PA, USA, 48-57.

[3] Moira Burke and Robert Kraut. 2008. Mopping up: Modeling Wikipedia Promotion Decisions. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (CSCW '08). ACM, New York, NY, USA, 27-36. DOI:http://dx.doi.org/10.1145/1460563.1460571

[4] Brian Butler, Elisabeth Joyce, and Jacqueline Pike. 2008. Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in Wikipedia. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08). ACM, New York, NY, USA, 1101-1110. DOI: https://doi.org/10.1145/1357054.1357227

[5] Ewa S. Callahan, and Susan C. Herring. Cultural bias in Wikipedia content on famous persons. *Journal of the Association for Information Science and Technology* 62, no. 10 (2011): 1899-1915. DOI: https://doi.org/10.1002/asi.21577

[6] Katherine Ehmann, Andrew Large, and Jamshid Beheshti. "Collaboration in context: Comparing article evolution among subject disciplines in Wikipedia." *First Monday* 13, no. 10 (2008). https://doi.org/10.5210/fm.v13i10.2217

[7] R. Stuart Geiger and Heather Ford. 2011. Participation in Wikipedia's article deletion processes. In Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11). ACM, New York, NY, USA, 201-202. DOI: https://doi.org/10.1145/2038558.2038593

[8] Aaron Halfaker, R. Stuart Geiger, Jonathan T. Morgan, and John Riedl. 2013. The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline. American Behavioral Scientist 57, 5 (May 2013), 664--688. DOI: http://dx.doi.org/10.1177/0002764212469365

[9] Brent Hecht and Darren Gergle. 2010. The tower of Babel meets web 2.0: user-generated content and its applications in a multilingual context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10). ACM, New York, NY, USA, 291-300. DOI: https://doi.org/10.1145/1753326.1753370

[10] Brent Hecht and Darren Gergle. 2009. Measuring self-focus bias in community-maintained knowledge repositories. In *Proceedings of the fourth international conference on Communities and technologies* (C&T '09). ACM, New York, NY, USA, 11-20. DOI: http://dx.doi.org/10.1145/1556460.1556463

[11] Aniket Kittur, Bongwon Suh, Bryan A. Pendleton, and Ed H. Chi. 2007. He says, she says: conflict and coordination in Wikipedia. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07). ACM, New York, NY, USA, 453- 462. DOI: https://doi.org/10.1145/1240624.1240698

[12] Aniket Kittur, Ed Chi, Bryan A. Pendleton, Bongwon Suh, and Todd Mytkowicz. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World wide web* 1, no. 2: 19.

[13] Aniket Kittur and Robert E. Kraut. (2008). Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (CSCW '08). ACM, New York, NY, USA, 37-46. DOI: http://dx.doi.org/10.1145/1460563.1460572

[14] Travis Kriplean, Ivan Beschastnikh, David W. McDonald, and Scott A. Golder. 2007. Community, consensus, coercion, control: CS*W or how policy mediates mass participation. In Proceedings of the 2007 international ACM conference on Supporting group work (GROUP '07). ACM, New York, NY, USA, 167-176. DOI: http://dx.doi.org/10.1145/1316624.1316648

[15] Shyong K. Lam, Jawed Karim, and John Riedl. 2010. The effects of group composition on decision quality in a social production community. In Proceedings of the 16th ACM international conference on Supporting group work (GROUP '10). ACM, New York, NY, USA, 55-64. DOI: http://dx.doi.org/10.1145/1880071.1880083

[16] David Laniado, Riccardo Tasso, Yana Volkovich, and Andreas Kaltenbrunner. (2011). When the Wikipedians Talk: Network and Tree structure of Wkipedia discussion pages." In *ICWSM*, pp. 177-184.

[17] Paul Laufer, Claudia Wagner, Fabian Flöck, and Markus Strohmaier. 2015. Mining cross-cultural relations from Wikipedia: A study of 31 European food cultures. In *Proceedings of the ACM Web Science Conference* (WebSci '15). ACM, New York, NY, USA, Article 3, 10 pages. DOI: https://doi.org/10.1145/2786451.2786452

[18] Paolo Massa and Federico Scrinzi. 2012. Manypedia: comparing language points of view of Wikipedia communities. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration* (WikiSym '12). ACM, New York, NY, USA, Article 21, 9 pages. DOI: http://dx.doi.org/10.1145/2462932.2462960

[19] Jonathan T. Morgan, Michael Gilbert, David W. McDonald, and Mark Zachry. 2013. Project talk: coordination work and group membership in WikiProjects. In *Proceedings of the 9th International Symposium on Open Collaboration* (WikiSym '13). ACM, New York, NY, USA, Article 3, 10 pages. DOI: http://dx.doi.org/10.1145/2491055.2491058

[20] Jonathan T. Morgan, Michael Gilbert, David W. McDonald, and Mark Zachry. 2014. Editing beyond articles: diversity & dynamics of teamwork in open collaborations. In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW '14). ACM, New York, NY, USA, 550-563. DOI: https://doi.org/10.1145/2531602.2531654

[21] Claudia Müller-Birn, Leonhard Dobusch, and James D. Herbsleb. 2013. Work-to-rule: the emergence of algorithmic governance in Wikipedia. In Proceedings of the 6th International Conference on Communities and Technologies (C&T '13). ACM, New York, NY, USA, 80-89. DOI: http://dx.doi.org.offcampus.lib.washington.edu/10.1145/2482991.2482999

[22] Keiichi Nemoto and Peter A. Gloor. "Analyzing cultural differences in collaborative innovation networks by analyzing editing behavior in different-language Wikipedias." *Procedia-Social and Behavioral Sciences* 26 (2011): 180-190. DOI: https://doi.org/10.1016/j.sbspro.2011.10.574

[23] Sabine Niederer, and José Van Dijck. "Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system." *New Media & Society* 12, no. 8 (2010): 1368-1387. DOI: https://doi.org/10.1177/1461444810365297

[24] Reid Priedhorsky, Jilin Chen, Shyong (Tony) K. Lam, Katherine Panciera, Loren Terveen, and John Riedl. 2007. Creating, destroying, and restoring value in Wikipedia. In *Proceedings of the 2007 international ACM conference on Supporting group work* (GROUP '07). ACM, New York, NY, USA, 259-268. DOI: http://dx.doi.org/10.1145/1316624.1316663

[25] Jodi Schneider, Alexandre Passant, John G. Breslin, (2010) 'A Content Analysis: How Wikipedia Talk Pages Are Used.' In WebSci2010, Web Science Conference. Raleigh, NC. http://hdl.handle.net/10379/2501

[26] Bongwon Suh, Gregorio Convertino, Ed H. Chi, and Peter Pirolli. 2009. The singularity is not near: slowing growth of Wikipedia. In Proceedings of the 5th International Symposium on Wikis and Open Collaboration (WikiSym '09). ACM, New York, NY, USA, , Article 8 , 10 pages. DOI: http://dx.doi.org/10.1145/1641309.1641322

[27] Nathan TeBlunthuis, Aaron Shaw, and Benjamin Mako Hill. 2018. Revisiting "The Rise and Decline" in a Population of Peer Production Projects. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, New York, NY, USA, Paper 355, 7 pages. DOI: https://doi.org/10.1145/3173574.3173929

[28] Fernanda B. Viégas, Martin Wattenberg, Jesse Kriss, and Frank Van Ham. "Talk before you type: Coordination in Wikipedia." In *System sciences, 2007. HICSS 2007. 40th annual Hawaii international conference on*, pp. 78-78. IEEE, 2007. DOI: https://dx.doi.org/ 10.1109/HICSS.2007.511

[29] Max L. Wilson, Ed H. Chi, Stuart Reeves, and David Coyle. 2014. RepliCHI: the workshop II. In CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14). ACM, New York, NY, USA, 33-36. DOI: https://doi.org/10.1145/2559206.2559233

[30] Max L. Wilson, Wendy Mackay, Ed Chi, Michael Bernstein, Dan Russell, and Harold Thimbleby. "RepliCHI-CHI should be replicating and validating results more: discuss." In *CHI'11 Extended Abstracts on Human Factors in*

Computing Systems, pp. 463-466. ACM, 2011. DOI: https://doi.org/10.1145/1979742.1979491

[31] Yasseri T., Spoerri A., Graham M., and Kertész J., The most controversial topics in Wikipedia: A multilingual and geographical analysis. In: Fichman P., Hara N., editors, Global Wikipedia: International and cross-cultural issues in online collaboration. Scarecrow Press (2014). DOI= http://dx.doi.org/10.2139/ssrn.2269392