2009

Cross-Cultural Analysis of the Wikipedia Community

Noriko Hara  
*Indiana University*, nhara@indiana.edu

Pnina Shachaf  
*Indiana University*, shachaf@indiana.edu

Follow this and additional works at: [http://aisel.aisnet.org/amcis2009](http://aisel.aisnet.org/amcis2009)

**Recommended Citation**

[http://aisel.aisnet.org/amcis2009/770](http://aisel.aisnet.org/amcis2009/770)
Cross-Cultural Analysis of the Wikipedia Community

Noriko Hara
School of Library & Information Science, Indiana University
nhara@indiana.edu

Pnina Shachaf
School of Library & Information Science, Indiana University
shachaf@indiana.edu

ABSTRACT

This paper reports a cross-cultural analysis of Wikipedia communities of practice (CoPs). First, this paper argues that Wikipedia communities can be analyzed and understood as CoPs. Second, the similarities and differences in norms of behaviors across three different languages (English, Hebrew, and Japanese) and on three types of discussion spaces (Talk, User Talk, and Wikipedia Talk) are identified. These are explained by Hofstede’s dimensions of cultural diversity, the size of the community, and the role of each discussion area. This paper expands the research on online CoPs, which have not performed in-depth examinations of cultural variations across multiple CoPs.

Keywords
Knowledge sharing, Wikipedia, communities of practice, cross-cultural

INTRODUCTION

The proliferation of the social Web and its participatory nature gives rise to many online communities, some of which undertake a common practice that links users together. In this environment, users often become “prosumers,” who are consumers and producers at the same time (Tapscott & Williams, 2008). These prosumers are sharing knowledge through massive collaborative efforts, for example, by writing encyclopedic articles on sites such as the Wikipedia, or by providing answers to questions posted on Q&A sites such as Yahoo! Answers. Various companies identified the potential utility of prosumers by soliciting product and research ideas from the prosumers, who are not employed by their organizations (Tapscott & Williams, 2008). Wikipedia users are engaged in knowledge sharing processes by implementing common practices that create and maintain shared identity. This type of knowledge sharing is pertinent to the idea of “communities of practice” (CoPs) (Lave & Wenger, 1991). The CoP framework offers “a lens by which we can focus our understanding of informal collaborative learning that occurs outside formal classrooms and training environments” (Hara, 2009, p. 2).

The literature about online CoPs is mostly descriptive in nature and focuses on what online CoPs are (e.g., Barab, MaKinster, & Scheckler, 2004; Gray, 2004; Schlager, M., Fusco, J., & Schank, 2002), or on illuminating the potential for online CoPs (Johnson, 2001; Sharratt & Usoro, 2003). Research on CoPs typically includes in-depth case studies of specific groups—math and science teachers (Barab, et al, 2004), healthcare professionals (Hara & Hew, 2007), lawyers (Wasko & Faraj, 2005), and employees in Caterpillar (Ardichvili, Page, & Wentling, 2003). A few analyses of multiple CoPs have developed a typology (Dubé et al., 2006) and examined factors that motivate or hinder participants from engaging in knowledge sharing in online environments (Hew & Hara, 2007). These studies enrich our understanding of online CoPs beyond a single case study. However, few cross-cultural analyses of CoPs have been conducted in the past. For instance, Pan and Leidner (2003) studied a knowledge management system that resides within an international organization to connect multiple CoPs, but they did not particularly focus on a cross-cultural analysis of CoPs. There is a need for analytical, conceptual, and comparative work regarding online CoPs that transcends common single case study approaches, which tend to be confined within organizational boundaries (Hara, Shachaf, & Storeger, forthcoming). In particular, as organizations become more global and deploy more cross-border CoPs than ever before, there is a need for cross-cultural analyses of CoPs.

The global nature of Wikipedia makes it an interesting case for such cross-cultural analyses; over 75% of Wikipedia is written in languages other than English. Despite the multi-lingual nature of Wikipedia (75,000 active members in more than 260 languages (Wikipedia: About)), prior studies are predisposed to investigate only the English version of Wikipedia, with a few exceptions (e.g., Pfeil, Zaphiris, & Ang, 2006). Research of non-English Wikipedia is needed to better understand the larger Wikipedia community. A cross-cultural analysis of Wikipedia as a CoP would increase our comprehension regarding questions about the impact of information and communication technology on cultural differences and would address questions such as: Do differences across various Wikipedia languages reflect the known cultural variations, or does Web 2.0 bridge these cultural differences? Are the similarities and differences across languages mediated by the shared practice of the
online CoP or by the online participatory platform? As a first step in addressing these questions, this paper focuses on the following research questions: How do Wikipedia norms of behavior vary across languages?; What are the (cultural) variations among Talk pages (Talk, Wikipedia Talk, and User Talk pages)?

BACKGROUND

Wikipedia as Communities of Practice

In this paper we use the following definition of CoPs: “groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis” (Wenger, Dermott, & Snyder, 2002, p. 4). Wenger (1998), in his seminal book, identified four characteristics of communities of practice: practice, community, meaningful learning, and identity. All four characteristics of online CoPs are prevalent within the Wikipedia community; the manifestation of each of these characteristics in Wikipedia is described next.

1. Practice: Wikipedia users share practice and are engaged in knowledge sharing as seen in discussion spaces (e.g., Stvilia, Twidale, Smith, Gasser, 2008). Users write original articles, edit and improve existing articles, provide quality assurance, fight against acts of vandalism, participate in policy setting, and are engaged in community building and maintenance activities. Sharing these practices fosters a sense of community among Wikipedia users.

2. Community: Wikipedia users developed a set of community norms and policies. Users are expected to adhere to these norms and the Wikipedia policies (e.g., Riehle, 2006; Shachaf & Hara, forthcoming). New members of the community are informed and encouraged to learn these norms of behaviors by seasoned users who have mastered processes, policies, and practices on the Wikipedia. Bryant, Forte, and Bruckman (2005) report that new users tend to focus on editing individual articles and later become motivated to contribute to the well-being of the Wikipedia community and the Wikipedia project as a whole.

3. Meaningful learning (learning in context): Wikipedia users learn to become Wikipedians (e.g., Bryant, et al., 2005); they learn how to behave, how to write, and how to be a member (in good standing) of the community. When examining how new Wikipedia users become Wikipedians, Bryant, Forte, and Bruckman, (2005) applied a concept of legitimate peripheral learning used in the communities of practice literature (Lave & Wenger, 1991). “Legitimate peripheral learning” is a term used to describe a mode of participation in which newcomers to a community peripherally participate in the practice. Though peripheral, the participation is legitimate in the sense that these new apprentices can observe other members, especially the more experienced, and learn how to become full members of the community. This learning was documented in Bryant, et al.’ (2005) study, focusing on users’ participation in Wikipedia over time, as the pattern and type of their contributions and motivations evolved over time.

4. Identity: Shared identity is partially formed around the share practice of the CoP. Anthony, Smith, & Williamson (forthcoming) contend that the Wikipedia user develops an identity as a Wikipedian. As shown among open source developers, reputations and group identity are one of the motivators to contribute to public goods. Due to the strong group identity on the Wikipedia community, contributions are flourishing.

Culture: Hofstede’s dimensions

Culture is defined as “the collective programming of the mind which distinguishes the members of one group or category of people from another” (Hofstede, 1991, p. 5). One way that scholars try to understand the nature of cultural differences is to understand a pattern composed of a combination of dimensions (Straub, Loch, Evaristo, Karahanna & Srite, 2002). While various such approaches to culture exist (e.g., Hall; 1976; 1983; Hofstede, 1991; Kluckhohn and Strodebeck 1961; Trompenaars & Hampden-Turner, 1998), Hofstede’s (1991) five dimensions are the most commonly utilized pattern in cross-cultural research. The dimensions include: low/high power distance, individualism/collectivism, masculinity/femininity, uncertainty avoidance, and long-term/short-term orientation. The fifth dimension was a later addition to his framework based on the contribution of Chinese scholars (Hofstede, 2008). Empirical studies that utilize the dimensional approaches, however, report inconsistencies in the context of groups that use technology. Most of the studies focus only on 3-4 dimensions, mostly on Hofstede’s (1991) dimensions (Mayers & Tan, 2002), and typically, they only focus on the individualism/collectivism dimension.

Power distance (high vs. low) refers to the distance among different people in varied levels of a hierarchy. It is “the extent to which the less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally” (Hofstede, 1991, p.28).
Individualism vs. Collectivism describes the orientation of an individual either first toward the group interest and then to his own, or first to his own and then toward the group interest. The group responsibility toward the individual is evident in a collectivist culture, yet almost non-existent in individualistic cultures where the individual is expected to look after himself. “Individualism pertains to societies in which the ties between individuals are loose: everyone is expected to look after himself or herself and his or her immediate family. Collectivism as its opposite pertains to societies in which people from birth onwards are integrated into strong, cohesive in groups, which throughout people’s lifetime continue to protect them in exchange for unquestioning loyalty” (Hofstede, 1991, p.51).

Masculinity vs. Femininity. Masculinity refers to more assertiveness and business-interest orientation, while femininity refers to taking care of quality of life and the individual need, which are external to the business. Hofstede (1991) suggests that “masculinity pertains to societies in which social gender roles are clearly distinct (i.e., men are supposed to be assertive, tough, and focused on material success, whereas women are supposed to be more modest, tender, and concerned with the quality of life); femininity pertains to societies in which social gender roles overlap (i.e., both men and women are supposed to be modest, tender, and concerned with the quality of life)” (p. 82).

Uncertainty avoidance (high vs. low) refers to the notion of taking or avoiding risks, and to the attitude toward uncertainty and ambiguity. Uncertainty avoidance is “[T]he extent to which the members of a culture feel threatened by uncertain or unknown situations. This feeling is, among other things, expressed through nervous stress and in a need for predictability: a need for written and unwritten rules” (Hofstede, 1991, p. 113).

Long-term vs. Short term dimension is a later addition to the four initial dimensions. Hofstede (1991) suggests that “Short term orientation [societies are characterized by] respect [for] tradition; respect for social and status obligations regardless of cost; social pressure to ‘keep up with the Joneses’ even if it means overspending; small saving quota, little money for investment; quick results expected; concern with possessing the truth...Long term orientation [societies are characterized by] adaptation of traditions to a modern context; respect for social and status obligations within limits; thrift, being sparing with resources; large savings quotes, funds available for investment; perseverance towards slow results; willingness to subordinate oneself for a purpose; concern with respecting the demand of virtue” (p. 173).

Research Gap and Questions
Prior research on CoPs is driven by case studies of CoPs, and research on Wikipedia primarily focuses on the English Wikipedia and tends to overlook the Wikipedias in languages other than English. This study examines the variation of norms among Wikipedia users in three languages. The study aims to answer the following research questions: How do Wikipedia norms of behavior vary across languages?; What are the (cultural) variations among Talk pages (Talk, Wikipedia Talk, and User Talk pages)?

METHODS
This section discusses the method of data collection and data analysis procedures. In an effort to answer the research questions, content analysis of 90 Wikipedia talk pages in three languages was conducted. These 90 pages were randomly selected from Wikipedia in three different languages—Hebrew, Japanese, and English; two languages are spoken as native language in Western cultures (English and Hebrew) and one in Eastern cultures (Japanese). These Wikipedias also vary in their sizes: two large Wikipedias, with over 100,000 articles each (English, with 2,580,417 articles, and Japanese, with 526,800 articles) and one smaller Wikipedia, with less than 100,000 articles (Hebrew, with 83,034 articles). Table 1 provides additional information on these three languages.

<table>
<thead>
<tr>
<th>Language</th>
<th>Rank by size</th>
<th>Articles</th>
<th>Share of Wikipedia</th>
<th>Number of Users</th>
<th>Articles per speaker</th>
<th>Edits per article</th>
<th>Ratio of users to administrators</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>1</td>
<td>2,580,417</td>
<td>22.69%</td>
<td>8,039,178</td>
<td>1,889</td>
<td>80.7</td>
<td>5,003</td>
</tr>
<tr>
<td>Japanese</td>
<td>5</td>
<td>526,800</td>
<td>4.67%</td>
<td>249,531</td>
<td>3,091</td>
<td>36.57</td>
<td>3,839</td>
</tr>
<tr>
<td>Hebrew</td>
<td>26</td>
<td>83,034</td>
<td>0.73%</td>
<td>65,000</td>
<td>8,599</td>
<td>63.22</td>
<td>1,300</td>
</tr>
</tbody>
</table>

Table 1. Descriptive information on four Wikipedias
Data Collection

Data collection was undertaken in two stages. The first stage was a pilot study on the English Wikipedia. The process of data collection from the English Wikipedia conducted in July 2007 was repeated on the other Wikipedias (Hebrew and Japanese) in September 2008.

Wikipedia’s massive information is saved under 20 name spaces; the biggest one is the “main” name space, which includes all encyclopedia articles (Viégas et al., 2007). Among these 20 name spaces, nine talk name spaces exist. Talk name spaces provide users with places in which they can discuss issues and communicate with each other. Talk name spaces are linked to other name spaces. For example, a talk page is associated with each Wikipedia article, which is called a “Talk” page. Likewise, a talk page is associated with each user page or Wikipedia policies page, which is called a “User Talk” page or a “Wikipedia Talk” page, respectively. In addition, an history page is linked to each Wikipedia page from all name spaces. These history pages provide links to older versions of that page as well as history pages for its respective talk pages; any of these pages may have an archive (in cases of intensive discussions on a specific talk page, some of the older discussions are archived). Viégas et al. (2007) identified the following nine talk name spaces: 1) Talk; 2) User Talk; 3) Wikipedia Talk; 4) Image Talk; 5) MediaWiki Talk; 6) Template Talk; 7) Help Talk; 8) Category Talk; and 9) Portal Talk. Among the nine, the top three most active talk name spaces, i.e., Talk, User Talk, and Wikipedia Talk (Viégas et al., 2007), were chosen for the analysis.

Sampling Process

To identify the norms of behaviors on the English Wikipedia, we randomly selected talk pages. Using the Wikipedia search capabilities, a list frame for the various talk pages was generated on July 18th, 2008 to use for the random sampling. Random sampling of Wikipedia pages is not a common practice; prior research on Wikipedia norms used purposeful sampling (e.g., Viégas et al., 2007). The sample in the first step of the pilot study included 30 pages, 10 pages from each of the three name spaces, Talk, User Talk, and Wikipedia Talk (Table 1). These 30 pages include more than 700 posts.

During the second step of data collection, the same sampling procedure was utilized to sample pages from the other two Wikipedias in September 2008. The sample at this step included 60 additional pages, from the two other languages (Hebrew
and Japanese). The final sample used in this study included 90 talk pages, 30 talk pages from each of the three Wikipedias and from three different name spaces, 10 from each type of talk page. In total, this includes over 2,500 Wikipedia messages.

**Data Analysis**

The coding scheme was developed from the ground up during the pilot study with the English Wikipedia, using interpretive content analysis; both authors coded a subset of the data and suggested a list of codes. These codes were discussed and grouped into broader categories. The final coding scheme includes ninety codes under 3 categories: writing norms, information sharing, and community well-being. This coding scheme was later used to code the pages and to compare norms of behavior among the three types of talk pages.

Each of the authors and an independent researcher coded the selected pages in one language, which was his/her native language. All three coders coded 10% of the data of the English Wikipedia to examine the level of inter-coder reliability (number of agreements, divided by the sum of the number of agreements plus number of disagreements). Inter-coder reliability was high: 93% between Hebrew and English Wikipedia, and 84% between the Japanese and English Wikipedia.

**Limitations**

One of the limitations of the study is an assumption that a Wikipedia in a specific language relates to a national culture of a specific country. The languages used in the Wikipedias, however, do not exactly correspond with a specific country. It is possible that the majority of the users of the Hebrew and Japanese Wikipedias are native speakers of the respective language and are residents of the respective countries (Israel and Japan). At the same time, the users who participate in the English Wikipedia are from many English speaking countries as well as other non-native English speakers from countries around the globe. Another limitation common to cross-cultural research is that multiple researchers, speakers of various languages, coded the data. Even though the inter-coder reliability on the English pages among all three coders was high, it is possible that some of the variations across languages are partially due to the differences among the coders.

**FINDINGS AND DISCUSSION**

Figure 1 presents the number of messages posted in each of the three Wikipedias. The English Wikipedia has the highest number of messages on each type of talk page. This occurs naturally as the English Wikipedia overwhelmingly has the largest contributions among the three (Table 1). In addition, the numbers of messages posted in Wikipedia Talk correspond with the size of Wikipedias (from the largest to the smallest: English, Japanese, and Hebrew).

Due to the space limit, the coding scheme was not included in this paper.
Findings also indicate that the content of messages varies across types of talk pages and that the pattern of variation is similar across all three languages (for example, postings about accuracy and quality of information are more frequently done on Talk pages than on User Talk pages, or Wikipedia Talk pages). The communication style in each language is consistent across talk pages but differs across languages. This pattern of variations correlates with Hofstede’s dimensions and the relative ranking of each country on the power distance index. The following sections describe these similarities and differences across cultures.

**Similar patterns across cultures**

Throughout all three languages, postings about accuracy and quality of information are more common on Talk pages than User talk or Wikipedia Talk, whereas more messages related to courtesy are found in User Talk than Talk or Wikipedia Talk pages. This indicates that messages posted on Talk and Wikipedia Talk pages tend to be more task-oriented. In contrast, messages posted on User Talk pages are more social in nature. This pattern is evident in all three languages.

**Quality and Accuracy**

Postings pertinent to quality and accuracy (facts accuracy, citations, and evaluations of contribution quality) in each of the three languages appear mostly on Talk pages. Figure 2 portrays percentages on each of the three talk pages. In all three languages there are more postings about quality and accuracy of information on Talk pages than on User Talk pages.

![Figure 2. Quality and Accuracy](image)

**Courtesey**

Courteous postings (for example, apologies, appreciation, and greetings) were posted more frequently on User Talk than any other type of talk page, across all three languages (see Figure 3). This occurs naturally as users tend to socialize on User Talk pages, not as much in other task-oriented talk pages.

**Variations across languages**

Across the three languages, variations in communication style exist on all three types of talk pages (Talk, User Talk, and Wikipedia Talk). These variations correlate with the existing Hofstede’s dimensions of cultural diversity, and in particular with the Power Distance dimension.
Variations of courtesy codes across languages (Figure 3) reveal that the Japanese pages have more messages that are courteous on each type of talk pages, than the English and the Hebrew pages. The rank order of courtiers on Hofstede’s (1991) Power Distance Index corresponds to the exhibition of courtesy on each of the languages respectively. People in various countries perceive power inequalities differently in the context of family, school, and work. In high power distance countries (i.e., Japan), more respect toward parents (by their children), teachers (by their students), and bosses (by their subordinates) is likely to be observed. According to Hofstede’s Power Distance Index, Japan is ranked 33rd and has the highest rate of courteous messages, the United States is 39th and has the second highest rate, and Israel is 52nd and has the lowest rate of courteous messages. Politeness is closely related to power distance (e.g., Daller & Yildiz, 2006). The higher the power distance in a society, the more it affects communication patterns; and the more distance individuals feel from others, the more likely they are to be polite.

![Figure 3. Courtesy](image)

**Indirect Writing Style**

Another variation in communication style across the three languages that corresponds with Hofstede’s Power Distance Index (1991) is in the use of direct/indirect writing style. Figure 4 illustrates the differences among the three languages in the use of indirect writing style. Among all the three name spaces, the Japanese Wikipedia has the most indirect messages, the English has the second most, and the Hebrew Wikipedia has the least percentage of messages written in an indirect style. The similar tendency was reported in Nelson (1997) that Americans are inclined to be more direct whereas Asian including Japanese tend to be more indirect in communication styles (see also Gudykunst & Ting Toomey, 1988).
CONCLUSION

This paper identified the Wikipedias as CoPs, and then examined three Wikipedia talk pages in three languages. It explained similarities and differences among the languages and talk pages by the roles of name spaces or Hofstede’s dimensions. In general, task-oriented postings, such as quality and accuracy, were found more frequently on Talk pages in all three languages than on other pages (Wikipedia Talk and User Talk). On the contrary, community well-being postings, such as courtesy, appeared mostly in User Talk pages in all three languages (compared with Talk pages and Wikipedia Talk pages). Thus, in any of these languages, Wikipedia users differentiate the use of each type of discussion area by posting messages of different natures and purposes. When differences among the three Wikipedias are found, these variations can be explained by Hofstede’s (1991) dimensions of cultural diversity, and the Power Distance index in particular. This study sheds light on how CoPs operate by analyzing norms of behaviors. In particular, the three Wikipedias that we examined provided examples of CoPs that exist in different cultural environments. Future research should expand the number of languages and examine whether the differences are statistically significant. As there are few studies of cross-cultural analysis about online CoPs, the Wikipedias provided nice test-beds to examine variations of norms of behaviors in different cultures. Future research should look into cross-cultural analysis of CoPs in various cultures.

REFERENCES


