How Open are Japanese Open Source Communities? A Comparison of Responses to Newcomer Inquiries in Japanese and English User Support Forums

Jonathan Lewis[1]

Graduate School of Social Sciences, Hitotsubashi University
cs00651@srv.cc.hit-u.ac.jp

1 Introduction

Researchers of open source software have noted that relatively small numbers of Japan programmers contribute core code to open source software projects that are not started domestically.[1] Others have noted low levels of Japanese participation in global developer communities [2, p.8]. From the perspective of Japanese IT policy this isolation carries a number of dangers: Japanese developers risk duplicating work being undertaken elsewhere; they miss chances to develop their skills working on cutting-edge projects; and they miss out on business opportunities that arise from having a good reputation and contacts in global communities. Open source software projects, meanwhile, miss out on the contributions of talented developers as well as on the wider adoption of the software in Japan as it benefits from the attention of developers acquainted with the needs of local users. While this article confines itself to Japan, the same issues arise concerning other East Asian countries.

A number of factors might account for this relative isolation of Japanese open source software communities: many Japanese users and developers have poor English skills and therefore prefer to communicate in Japanese; time zone differences make real-time communication with US- or Europe-based developers difficult; and developers may have so much work available in Japan that they lack economic incentives to engage in core code development or in international developer communities. This article proposes another factor: that domestic Japanese user communities fulfill the needs of most of their members sufficiently well that members do not need to participate in English-medium mailing lists and forums. As a result, Japanese users will not find themselves drawn along the user-developer continuum that gradually turns newcomers into core contributors.

In order to test this proposal, I examine archives of posts to Japanese- and English-language mailing lists and forums for three open source Web content

[1] For example, Weber analyzes the numbers of Linux code authors as a percentage of their countries' populations, and finds a score for Japan much lower than that of most Western countries and even the Republic of China [1, p.68].

management systems: Geeklog, Plone and Wordpress. All three systems are widely used in Japan as well as around the world. In particular I address two questions:

1. What is the response rate and time to newcomer inquiries on Japanese- and English-language mailing lists and forums? I find that while newcomers to Japanese-language mailing lists and forums can expect to wait at least twice as long for a response as posters to English-language lists and forums, nevertheless they can expect a reply within a working day, and moreover are less likely to be ignored than those who post inquiries to English-language lists and forums.

2. Do peaks of traffic on Japanese-language mailing lists and forums coincide with quiet periods on English-language mailing lists and forums? The motivation for this question is that Tang et al [3] problematize the slower responses to newcomer inquiries from time zones outside the European and North American hubs of open source development. In my case study, an hour-by-hour breakdown of response rates and times is difficult due to insufficient numbers of posts in Japanese. Nevertheless, the findings do not suggest that response times of Japanese-language forums are generally shorter than those of English-language forums during Japanese business hours.

Organization of the paper. The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 presents my research questions and the result of the case study. Section 4 discusses the threats to validity of my work. Finally, Section 5 concludes the paper.

2 Related Work

The relevant existing research falls into two areas: research on the connections between Japanese and international open source communities, and research on the analysis of mailing lists.

The most significant survey of the connections between Japanese and international open source communities is the FLOSS-Japan survey of developers carried out in 2003.[4] Asked about their participation in global communities, 62.3% of respondents reported their involvement being mainly domestic, 16.1% mainly global, and 21.6% both domestic and global. The report's authors commented that 40% involvement in global communities is high for a hobby activity, and that global involvement may be one motivation for involvement in OSS.

The survey also asked respondents to rate their English ability. As one might expect, 83.3% of respondents participating only in domestic communities say their English is poor. But the report argues that "since a certain number of developers participate in the global community even though their English language ability is poor, it appears that participation is possible if one has sufficient courage and perseverance."

The FLOSSWORLD survey carried out between 2005 and 2007 does not cover Japan, but mentions in passing that "Software developers in developing countries (and some other reasons [*sic*] such as Korea and Japan) often face the problem that they do not integrate in the global FLOSS Communities because of language barriers."[2, p.8] The glass of Japanese participation in global open source communities, in other words, is half empty seen from outside Japan but half full from a domestic perspective.

While surveys of developers are invaluable to our understanding of links between Japanese and global open source communities, it is also necessary to examine the day-to-day reality of differences and links between the communities. This article starts to fill that gap.

The second area of relevant work is research on mailing lists. Given that most open source developers communicate on mailing lists, and that the archives of those mailing lists tend to be publicly available, it is not surprising that researchers have analyzed mailing lists from a number of perspectives including social structure and the code review process; [5] offers a good summary of this research. Some of this work has concentrated on trying to ascertain the geographical location of message posters from information such as email addresses, timezone information, and the IP addresses of mail servers; [6] represents the state of the art in this area.

The most relevant research on mailing lists for this article is Tang et al.'s analysis of global participation in two long-established open source projects [3]. In order to measure how open communities are to those from countries without a strong tradition in open source development, they analyze the response rate and time to inquiries by newcomers to the developers' (English-language) mailing lists for each hour of the day. They find that "the speed and rate of responding to inquiries by newcomers depend on the country of the email sender and the time that the email was posted to the mailing list." Their article offers an excellent template for analyzing the socio-geographic aspects of open source mailing lists, and the case study that follows makes use of their work. Their article, however, confines itself to an analysis of communications in English. In order to gain a fuller understanding of the global structures of open source communities, we also need to analyze communication between open source users and developers taking place in other languages.

Research on open source mailing lists has also concentrated on developers, i.e. those planning and carrying out the production of code, rather than users, i.e. those implementing the code in various real-world projects. There is, of course, no clear distinction between the two categories, as developers are almost always also users; and every open source project encourages casual users to involve themselves more deeply. Nevertheless, on an operational level, projects develop separate channels for dealing with basic questions about how to use the code and for discussing how to improve it. The English-medium Plone-Users mailing list, for example, covers development (of websites) *with* Plone, while the Plone-Dev list covers development *of* Plone. For subscribers to the Plone-Users list, it is a natural progression to subscribe to Plone-Dev when they

become interested in improving the project's code. They already know many of the Plone-Dev participants from Plone-Users and have a sense of how the community works and talks. A keen Japanese user who has got to the know the project in the Japanese Plone user mailing list, however, is not situated on this user-developer continuum and faces a major social, as well as linguistic, hurdle if he or she is suddenly to debate code improvements on the Plone-Dev list.

One technical consequence of this emphasis on developers rather than users is that researchers tend to concentrate on the analysis of mailing lists rather than other means of online communication. In two of the three projects analyzed below, web forums rather than mailing lists are used for communication between users. From the point of view of the researcher, web forums have some pros and cons compared with mailing lists. On the positive side, with forums every topic is clearly listed on one page; this is better than mailing lists, where it is necessary to reconstruct threads but not always possible to do so totally accurately. On the negative side, forums make it more difficult to determine the geographical location of a participant; unlike mailing lists, no time zone or IP address information is available, and only a minority of participants register a geographical location in their profile.

3 Case Study

The English and Japanese user support forums of three open source content management systems were selected. Each of the systems (Geeklog, Plone and WordPress) is mature, widely adopted in Japan and overseas, and has a stable population of end-users (who build sites by themselves), integrators (who use the systems to build websites for customers) and developers. Of course, the three systems themselves differ considerably in their internal workings and in their target users. Details of the user forums and mailing lists analyzed are given in Table 1.

Table 2 shows the results of the analysis. For each of the projects, many more messages have been posted in English than in Japanese. This is to be expected, and is also partly attributable to the fact that the English-language forums cover a longer period, from just under two years longer in the case of WordPress to 4 1/2 years longer in the case of Geeklog. Nevertheless, we do notice that the proportion of Japanese to English messages is much lower for Plone than for Geeklog and WordPress. This might indicate that, compared to the other two CMSs, Plone is less widely used in Japan than globally. This suggests a new avenue of research: comparing the Japanese diffusion of different open source projects through the proportion of Japanese to English messages on user forums.

Overall response rates and median response times to newcomer inquiries were calculated for each project. I used Tang et al's [3] definition of "newcomer" as someone who has not posted more than 20 times to a mailing list. In each project, response rates were high for both Japanese and English, al-

Table 1. Details of forums and mailing lists analyzed

project	language	forums/mailing lists analyzed
Geeklog	English	Installation, Plugin Help, General Help, Themes http://www.geeklog.net/forum/index.php 10 September 2001 - 5 March 2010
	Japanese	Installation, Themes, Language Files, Individual Functions, Plugins, Hacks, Mobile Sites, Settings, Bug Fixes http://www.geeklog.jp/forum/index.php 4 May 2006 - 6 March 2010
Plone	English	Plone Users http://news.gmane.org/gmane.comp.web.zope.plone.user 3 Oct 2002 - 27 February 2010
	Japanese	Plone-Users http://ml.plone.jp/mailman/listinfo/plone-users 22 May 2005 - 24 February 2010
WordPress	English	Installation (part), How-To and Troubleshooting (part) http://wordpress.org/support/ 24 May 2006 - 8 March 2010
	Japanese	Installation, How-To and Troubleshooting, Plugins, Themes, Bug Reports and Suggestions http://ja.forums.wordpress.org/ 26 March 2008 - 5 March 2010

Table 2. Results of forum and mailing list analysis

	Geeklog		Plone		WordPress	
	English	Japanese	English	Japanese	English	Japanese
Total messages	61,442	7,086	105,222	2,369	84,786	13,579
Total threads/topics	14,754	1,928	33,243	663	21,077	3,244
Threads started by newcomers	10,359	948	15,180	393	20,387	2,503
Response rate to newcomer threads	87.6%	91.5%	64.7%	77.8%	85.3%	89.0%
Median response time to newcomer threads (hours)	2.00	2.25	3.01	6.69	1.08	4.38

though significantly higher for Geeklog and WordPress than for Plone. More cases would have to be studied to know whether this difference is a result of the medium or of the particular communities. For example, it may be that a number of newcomer messages to mailing lists are of a one-way nature, such as announcements, or emails that request private replies. Such factors would result in a lower response rate than on web forums, regardless of how open or helpful the community.

In each of the projects, response rates are higher for Japanese than for English forums, although only by a small margin in the cases of Geeklog and WordPress. This is strong evidence for the argument that Japanese open source communities are able to fulfill users' needs.

Next, once again following Tang et al [3], I calculated hour-by-hour median response times. Might Japanese forums respond quicker during times when English forums are sluggish? In other words, are Japanese users resorting to vernacular communication only because the global community is asleep?

The results are shown in Figures 1, 2 and 3. The problem here is that the number of newcomer inquiries to Japanese-language forums is too low for reliable analysis. The results for Plone, in particular, are of little use, with no inquiries at all in the early hours of the morning and massive distortion from outlying results (see Figure 1). To the extent that results for Geeklog's Japanese forum are reliable, response times are faster than for its English forum at certain times of day, but not during Japanese business hours. The most reliable results are for WordPress, and Figure 3 neatly shows (a) that response times in both English and Japanese tend to rise and fall at the same time of day from the point of view of a Japanese user, and (b) that response times for English forums are always better than those for Japanese forums. So we have no evidence that time zone-induced slow response time in the global community is boosting Japanese participation in vernacular forums.

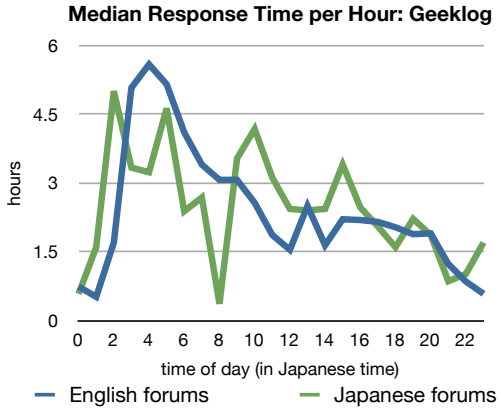

Fig. 1. Response times per hour: Geeklog

4 Threats to Validity

In order to be more confident of my conclusions, I need to increase the number of cases. The projects covered should include those that use web forums and those that rely on mailing lists. One problem, however, is that many Japanese communities simply do not generate enough messages to permit reliable hour-by-hour analysis.

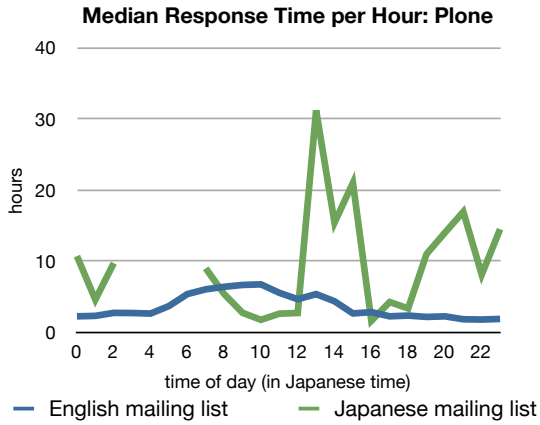

Fig. 2. Response times per hour: Plone

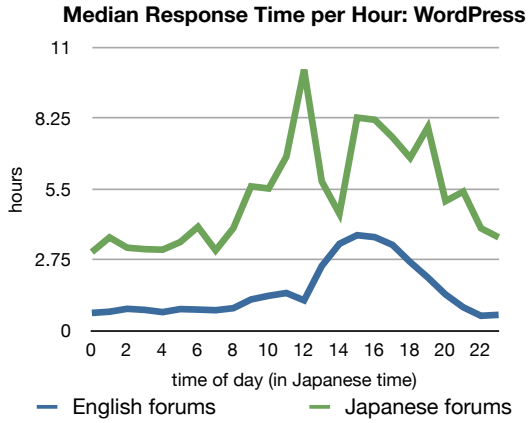

Fig. 3. Response times per hour: WordPress

It is also possible that web content management systems, which have a wide user base of less technically sophisticated users, might have atypically structured communities in terms of users and developers; in particular, if there are large user communities, it is more likely that local (non-English) communities will grow up in relative isolation from the English-medium developer communities. On the other hand, it might also be argued that these projects with large end-user communities have exceptional value as a potential pool of open source developers and should therefore be studied even if they are atypical.

5 Conclusion

This paper has argued that Japanese open source communities are highly open and remarkably responsive to newcomers. This openness represents a great effort on the part of a small number of active members. It is, of course, very good for users and almost certainly accelerates the diffusion of open source software within Japan. On the other hand this very openness, by meeting the needs of most users in a Japanese-medium environment, might serve as a barrier to the progression of participants from new user to core developer which can be observed in English-medium communities.

Each project has its own characteristics and history, so there can be no one-size-fits-all solution to the problem of community isolation. However, based on the author's experience in the Japanese Plone community since 2004, the efforts of a small number of community members to participate in international meetings, to bring key developers to Japan and to encourage fellow members to communicate with the international community can, in the long term, bear fruit in terms of Japanese developers starting to participate in the core project.

References

1. Weber, S.: The success of open source. Harvard University Press, Cambridge, MA (2004)
2. David, P.A., Ghosh, R.A., Glott, R., Gonzalez-Barahona, J.M., Shapiro, J.: FLOSSWorld. Final research report and policy impact, United National University-MERIT, Maastricht, Netherlands (June 2007)
3. Tang, R., Hassan, A.E., Zou, Y.: A case study on the impact of global participation on mailing lists communications of open source projects. In: Proceedings of the 3rd International Workshop on Knowledge Collaboration in Software Development (KCSD), Kyoto, Japan (November 2009)
4. Mitsubishi Research Institute: Free/libre/open source software Japanese developers online survey (FLOSS-JP). English version of survey results., MRI, Tokyo, Japan (2004)
5. Shibab, E., Bettenburg, N., Adams, B., Hassan, A.E.: On the central role of mailing lists in open source projects: An exploratory study. In: Proceedings of the 3rd International Workshop on Knowledge Collaboration in Software Development (KCSD), Kyoto, Japan (November 2009)
6. Tang, R., Hassan, A.E., Zou, Y.: Techniques for identifying the country origin of mailing list participants. In: 16th Working Conference on Reverse Engineering. (2009)