

# A sociolinguistic consideration of free and open source software teams as virtual communities of practice

By Jeremy Freudberg

---

The free and open source software (FOSS) development world is one in which people regardless of background come together to share the common goal of creating non-proprietary software, and intend to do so in the most inviting and open way feasible. There is a unity centered around the common goal of freedom. This notion permeates all throughout the sub-communities which make up this larger community, and it cannot be avoided. The Free Software Foundation has gone on to claim that this is really a “political idealism” (Stallman, 2014) which may further explain why these views must be so actively perpetuated. Most members of these communities further hold beliefs that “[w]ith proprietary software, the program controls the users, and some other entity (the developer or ‘owner’) controls the program. So the proprietary program gives its developer power over its users. That is unjust in itself, and tempts the developer to mistreat the users in other ways” (Stallman, 2017). This quasi-fear allows the identity of the FOSS community to be firmly grounded. There is a general sentiment that in order for the community to exist, it must exist in stark opposition to the alternative of proprietary software, given the size of the rift between proprietariness and freedom. The competing software concepts will always be ‘at odds’ with one another, so this must be reflected in the behavior of the community.

These sentiments and behaviors can be observed in the contributors to the OpenStack software, which is a large FOSS project offering cloud computing capabilities, and which faces competition from proprietary cloud solutions. In this paper, the subproject OpenStack Sahara will be explored specifically. Sahara offers functionality beyond the base cloud capabilities, and is therefore driven by a smaller team and essentially at the fringes of its parent community. Sahara’s constituent members, therefore, must secure their identity in the FOSS world as staunchly as possible: in order to compensate for the smaller splash that they may make, they must make a larger investment in the philosophy of the community. Additionally, this author is a major contributor to OpenStack Sahara.

FOSS groups do not limit themselves to members of a single background, and Sahara is no exception. In fact, the community of practice of OpenStack Sahara is geographically dispersed across four continents, with contributors having ranged in age from young adulthood to late middle-age. Additionally, one can observe that the community has contained males, females, and at least one transgender individual. Therefore, what endows

this community a common identity truly is, in this case, the aforementioned philosophical goals, not any physical or demographic cohabitation. In fact, what allows the philosophy and principles of the community to bind the group so tightly is that notable dispersion. If the group were to be physically localized, or otherwise demographically close, then there would not be such a need to cultivate this additional principled identity. This type of community of practice, one whose division is the source of its unity, lies in contrast to traditional communities of practice.

Growing out of the above discussion, it could be argued that because this particular community of practice is a construction which exists almost entirely within the realm of the Internet, we should not expect it to exhibit the same properties as offline communities of practice. The legitimacy of virtual communities of practice, however, has already been well-established by other scholars.

First, one may consider some parallels between the virtual community of practice (VCoP) in this paper, OpenStack Sahara, and one which is much more established, Wikipedia. Contributors to Wikipedia have been identified as forming a VCoP while also being dispersed (Bryant, 2005). This is equivalent to the composition of Sahara contributors. Moreover, a crucial similarity between the established VCoP of Wikipedia and the potential VCoP of Sahara is that members are once again united by philosophy and principle. It has been further noted that “many Wikipedians perceive their work as contributing to a greater good, offering knowledge to the world at large. When asked why they contribute to the Wikipedia, many Wikipedians recognized the project’s overarching goals, the appeal of community, and perceived contributions to society” (Bryant, 2005, p. 4). It is clear that the common and principled sense of purpose is enough to unify the random collection of users into a true community of practice.

Another study formalized this theory of what is necessary to bind a VCoP, by exploring quantitatively what is necessary for the “success”, i.e. endurance, of a VCoP. Crucially, a study which examined the characteristics of successfully- and unsuccessfully-launched VCoPs has hypothesized that “not all virtual communities of practice are created equal” (Dubé et al., 2005, p. 148). Therefore, from that inequality, it is possible that in some cases being Internet-based may have a negative effect on the endurance of the VCoP, but in other cases this effect may not be observed. The study later clarifies this inequality, in offering a conclusion that “the structuring characteristics of a VCoP, especially the environment and relevance, are of utmost importance if an organization wants to increase the likelihood of a VCoP succeeding” (Dubé et al., 2005, p. 162). The takeaway here in terms of whether or not an online community will be a successful, enduring, and properly-formed VCoP is that key notion of relevance. The study further explains that notion, that “members feel that their contributions are worthwhile” (Dubé et al., 2005, p. 162). FOSS communities meet this criterion. Since, as mentioned earlier, the behavior within a FOSS community revolves almost entirely around promoting free-

software principles, contributions will no-doubt be worthwhile as they help to sustain those principles. Therefore, it can be accurately claimed that FOSS communities are robust– in other words, enduring and well-formed– VCoPs and consequently analogous in that robustness to a traditional offline community of practice.

It is important to note that despite the existing literature on VCoPs, not much about these communities has been examined through a sociolinguistic, or even simply linguistic lens. Bucholtz (1999) set a strong precedent for the applicability of the concept of “community of practice” to sociolinguistic inquiry, so an extension of that notion to VCoPs is a fair next step. Building towards that step, there has already been at a basic level some probing into the role of language in VCoPs, thereby skimming the surface of potential linguistic observations to be made. A 2012 study of the contributions made within the FOSS VCoP known as Mozilla examined the properties of the online discourse of the periphery members versus the central members. The results found mainly illuminated how language reflected how “the contribution of the periphery in Open Source communities consists mainly in declaring problems, asking for instructions, or for instructions concerning it, without intervening in a significant way in its solving” (Masmoudi et al., 2012, p. 327). These observations only dealt with quantitative formation of identity within the community, fitting squarely with Eckert (2012)’s notion of the “first-wave” of variation study. Explorations of externally-facing stylistic identity practices in the manner of Eckert’s “third wave” were not present.

Another study did manage to capture that element of externality, but still failed to make any claims about the formation of identity. In one study, some linguistic perspective was offered onto which kinds of online communications done by newcomers to virtual communities would receive replies, and which would be ignored. The principal finding of the study revolved around the idea of lurking, and how discourse that makes reference to that notion is better received: “referencing lurking, specifically, demonstrates that the author has committed time and effort... These forms of introductions signal a desire to belong” (Burke et al., 2007, p. 24). First, from this observation it again follows that the community of practice here is tightly bound, since in this case admission must be earned. It also demonstrates that online communities of practice are capable of maintaining norms, which since Gumperz (1964) have been consistently identified as an integral component of any speech community. In this case, lurking is perceived positively, as it means that the lurker has taken the time to observe those norms. Crucially, this is also certainly a discussion of style, and of external relations between communities, but there is no claim about using language to establish the identity of the community of practice. At this point, scholars have gotten increasingly closer to uncovering the role of sociolinguistic style and other third-wave notions in the identities of VCoPs but have still yet to reach it. This paper pioneers such investigations.

Before continuing onward to the following discussion of linguistic phenomena, a crucial fact to recognize is that some linguistic norms are dictated by the medium rather than the identity of the group. One example of a medium-driven norm is the widespread use of vocatives, as opposed to no direct address at all or pronouns. While this trend could be considered quite interesting in a non-virtual, real-life speech context, it is here a product of the fact that much of the interaction within this community is done on Internet Relay Chat (IRC). The heavy use of vocatives owes itself to practicality: members will receive a ‘ping’ upon the mentioning of their names but the natural flow a sentence will not be disrupted. This is worth noting as a prominent feature of the discourse, but it must not be identified as a stylistic consequence of the identity of any VCoP.

Returning to the discussion of the community of OpenStack Sahara, one prevalent linguistic phenomenon is the use of hedging in the community’s online discourse. Hedging is a catch-all term in pragmatics which describes the use of tacked-on words to lessen the impact or authority of an utterance. Consider the following examples from the Sahara team public chatroom: (in which the bolded utterances are of note)

1.  
`<elmiko>` so, in answer to the question from the email  
`<elmiko>` **i would say** prioritize the other work before microversions  
`<elmiko>` **you will need to** have a solid foundation to build the microversion stuff on top of (ie making sure you can support older code paths), but **so long as you are aware of that i think** you can wait until the new api looks stable to add them  
`<tellesnobrega>` that makes total sense
2.  
`<jeremyfreudberg>` tellesnobrega - **just want to remind you** about investigating the pig-oozie thing  
`<tellesnobrega>` jeremyfreudberg, right, can you remind me what am I investigating, can't remember now  
`<jeremyfreudberg>` tellesnobrega, **lol, sure**, basically the issue is pig libs are only installed when oozie is also installed. also those pig libs live in HDFS (oozie sharelib) so regular user can't even access them without oozie workflow

In the two above scenarios, the primary manifestation of hedging is that no absolutely direct suggestions or demands are made. In (1), the bolded phrases could simply be omitted yielding a much more blunt discourse. In (2), the bolded phrases aim to diffuse any sense of nagging regarding the reminder. Overall, the hedging phenomenon found in this discourse grows out of a sense that identity of this community of practice is founded on the need to be inoffensive, open, and kind– dissimilar from the perceived impersonal nature of proprietary software. In essence, these norms grow out of the need to reject the (alleged) negative values of the proprietary software world. This use of style can be considered a negative identity practice, as it revolves around the purposeful use of language features which would distance (at least in their own perception) members of that community from members of other communities.

Similar to the use of hedging is the use of ‘offering’ constructions. In place of constructions such as “I will”, which is a direct statement of intention, softer constructions are favored by Sahara contributors. Note the following excerpted statements, which are each prefixed by the form “let me”:

`<tellesnobrega>` **let me** take a look [12:56]

<tellesnobrega> let me just see if we can configure it before the files goes to the cluster [12:59]  
 <tellesnobrega> let me find a better example [13:04]  
 <tellesnobrega> let me look [13:17]  
 <tellesnobrega> let me see if I can get some examples [13:18]

First, it is worth noting that these statements were produced by a central member of the Sahara team, and directed towards a more casual contributor to the software. This is a natural context in which to assert the identity of the community, as the situation is somewhat external-facing. The construction is also used consistently and often in this extended discourse, so it is worth exploring as a marker of style. This ‘offering’ construction carries the meaning of humble volunteering rather than an absolute course of action. Its use is therefore highly compatible with the identity which this FOSS VCoP aims to cultivate. Namely, this particular instance of style promotes a people-first nature and distribution of power, an assertion of attributes which FOSS communities claim themselves to have, and which they claim proprietary software communities to lack. This use of style is again a negative identity practice, in that it distances this community from others, by an effect similar to the aforementioned use of hedging.

Another negative identity practice is that emoticons– established as a full-fledged feature of language by Dresner and Herring (2010)– and other emotional language can be found in the Sahara chatroom. Observe the following excerpts:

1.
 

<shuyingya\_m> Btw, vgridnev. My visa application has been rejected last Friday. I need to go to the American consulate after Chinese spring festival. I am not sure now whether I can take part at PTG.  
 <tellesnobrega> shuyingya\_m, that sucks :(  
 <egafford> shuyingya\_m: D:
2.
 

<jeremyfreudberg> lucky i only live one mile from the summit this time :p  
 [...]  
 <esikachev> jeremyfreudberg: i only live 2000 miles from the summit Ж)

The impetus behind this use of emotional language and emoticons is a conscious desire to distance the identity of those in FOSS community from the supposed monolithic, bland, impersonal, and unsympathetic world of proprietary software. (Recall Stallman’s definitions, in which the proprietary lacks the democratic– and therefore personal– aspects of the free and open.) The positioning of the FOSS community is, by this, a strange one, as seemingly positive aspects of language use are actually used as negative identity practices due to a (perhaps skewed) perception of the dominant norms held by those in the software realm at-large.

Ordinarily, FOSS communities of practice tend to think of themselves as the ‘rebels’ of software; this is why employment of mainly negative identity practices (linguistic or otherwise) has been observed so frequently. However, these communities are not entirely ‘underground’, so the employment of at least some positive identity practices may serve them well. Namely, members are incentivized to maintain, to some extent, the general attractiveness of their community as perceived by the surrounding software development sphere, regardless of any rebellious undertones. Data from the

Sahara chatroom can again be used as evidence. This time the style observed is that despite being founded on the principles of rebellion and revolution, FOSS communities of practice still put the discussion of software itself as a priority in their discourse. In order to assimilate with the greater community to some degree, the principles of the 'revolution' are not constantly dwelled upon or repeated. In fact within the Sahara chatroom it is incredibly rare to see members talk of high-level ideals from their pedestal. This is one of the few positive (conformist) identity practices of FOSS communities.

Much has been discussed in this paper, from the principles binding the free and open source software community, to the legitimacy of virtual communities of practice, and the role of third-wave sociolinguistic style in these communities. There is still much more to further explore. Potential topics of further investigation include VCoPs larger than OpenStack Sahara, FOSS communities which are also geographically or demographically close, and sociolinguistic norms in avenues of communication beyond the chatroom (either in actual spoken speech, in published materials, or in "code review" as seen in the 2012 Mozilla study). A key takeaway is that FOSS may be a breeding ground for sociolinguistic style just as any non-virtual community may be— especially since FOSS may be more actively engaged in rebelliousness and strictly focused on revolutionary identity more than many other communities of practice, virtual or otherwise. This new approach to understanding the community of practice will prove increasingly useful in the years to come as the goals of free software become more important and humanity becomes increasingly digital.

## REFERENCES

- Bryant, S. L., Forte, A., & Bruckman, A. (2005). Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, 1-10.  
doi:10.1145/1099203.1099205
- Bucholtz, M. (1999). "Why be normal?": Language and identity practices in a community of nerd girls. *Language in society*, 28(2), 203-223.  
doi:10.1017/S0047404599002043
- Burke, M., Joyce, E., Kim, T., Anand, V., & Kraut, R. (2007). Introductions and requests: Rhetorical strategies that elicit response in online communities. *Communities and Technologies*, 3, 21-39. doi:10.1007/978-1-84628-905-7\_2
- Dubé, L., Bourhis, A., & Jacob, R. (2005). The impact of structuring characteristics on the launching of virtual communities of practice. *Journal of Organizational Change Management*, 18(2), 145-166. doi:10.1108/09534810510589570
- Dresner, E. and Herring, S. C. (2010). Functions of the Nonverbal in CMC: Emoticons and Illocutionary Force. *Communication Theory*, 20, 249-268.  
doi:10.1111/j.1468-2885.2010.01362.x
- Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual review of Anthropology*, 41, 87-100.  
doi:10.1146/annurev-anthro-092611-145828
- Gumperz, J. (1964). Linguistic and Social Interaction in Two Communities. *Language, Culture and Society*, 14, 283-299. doi:10.1525/aa.1964.66.suppl\_3.02a00100
- Masmoudi H., Boughzala I. (2012) A Linguistic Analysis on How Contributors Solve Software Problems in a Distributed Context. *Open Source Systems: Long-Term Sustainability*, 322-330. doi:10.1007/978-3-642-33442-9\_29
- Stallman, R. (2014, June 24). *Motives For Writing Free Software*. Retrieved from <https://www.gnu.org/philosophy/fs-motives.en.html>.

Stallman, R. (2017, April 10). *Free Software Is Even More Important Now*. Retrieved from <https://www.gnu.org/philosophy/free-software-even-more-important.en.html>.

---

The relevant chat logs can be found at the following locations:

- <http://eavesdrop.openstack.org/meetings/sahara/2017/sahara.2017-04-06-14.00.log.html>
- <http://eavesdrop.openstack.org/meetings/sahara/2017/sahara.2017-03-23-14.00.log.html>
- <http://eavesdrop.openstack.org/irclogs/%23openstack-sahara/%23openstack-sahara.2018-08-16.log.html>
- <http://eavesdrop.openstack.org/meetings/sahara/2017/sahara.2017-01-26-14.00.log.html>
- <http://eavesdrop.openstack.org/meetings/sahara/2017/sahara.2017-03-16-18.03.log.html>
- Please contact the author for additional chat logs from his community.