Modes of Participation, Identity Creation, and Group Maintenance in Different Technical Discourse Communities

Society for the Social Studies of Science
Washington, DC (Crystal City)

stuart mawler
2009/10/30
What is this all about?

- Normatively identity-neutral interactions “hide” identity work
- Identity-oriented purposes are “hidden” in form and function of discourse events
- Community constructs combine
  - Discourse Communities
  - Epistemic Cultures
  - Communities of Practice
- Newsgroup analysis will illuminate these themes
  - Research still unfolding
Newsgroup Backgrounder

- Newsgroups are subsections of the Usenet system
  - Created in 1979 at Duke University
- Users read and submit public messages (“posts”) to newsgroups
- Discussions are threaded
  - Largely driven by news reader client software
- Still available and see high-traffic use
  - Via portals like Google and Yahoo
Compare newsgroup discourse for C++ and Lisp

Analyze all posts for 15 days: 7/1/2008-7/15/2008
- comp.lang.c++ (1440 messages in 238 threads)
- comp.lang.lisp (649 messages in 95 threads)

Why these newsgroups?
- Significantly different focus of each group
- Both are high-traffic (in top ten within comp.lang.* branch)
- C++: Largest single subscriber base in comp.lang.*
What Do Newsgroups Do?

- “Official” or Normative purposes
  - Solve intractable technical questions
  - “Be there for the newbie”

- Identity-oriented purposes
  - Determine what counts as knowledge
  - Create and maintain group mythologies
  - Pass-on shared history
  - Validate and maintain practices
  - Establish individual credibility
  - Maintain and strengthen boundaries with other languages
Technical Foundations... Mostly

- **C++**
  - High-level language, Object-oriented
  - General purpose
    - Commercial
  - Created by Bjarne Stroustrup at Bell Labs in 1979
  - ANSI/ISO standards maintained
    - Strongly “defended” or “protected”

- **Lisp**
  - Early high-level programming language
  - Central to classic AI research
    - Academic
  - Created by John McCarthy at MIT in 1958
  - Many sub-dialects, with two dominant
    - Common Lisp, created 1984
      - ANSI standard established 1994
    - Scheme, created 1970s
      - IEEE standard established 1990
Newsgroup “Style”

- Both unmoderated – community and individual policing
- Both dominated by small group
  - C++ top 5 = 24%; Lisp top 5 = 26%
  - C++ dominated by one guy (James) at 10.83%
- comp.lang.c++
  - Standards for interaction published
  - Regular reference to “the standard”
  - Regular answer writers rarely start threads
- comp.lang.lisp
  - No official standard of interaction
  - Less policing of off-topic posts
  - References to standards more diffuse
  - More “professionals” initiate posts
- Differences are *qualitative*, rather than *quantitative*
## Piles of Statistics

<table>
<thead>
<tr>
<th>Category</th>
<th>C++</th>
<th>Lisp</th>
</tr>
</thead>
<tbody>
<tr>
<td>total posts</td>
<td>1440</td>
<td>649</td>
</tr>
<tr>
<td>total threads</td>
<td>238</td>
<td>95</td>
</tr>
<tr>
<td>avg posts / thread</td>
<td></td>
<td></td>
</tr>
<tr>
<td>longest thread</td>
<td>46</td>
<td>117</td>
</tr>
<tr>
<td>top three threads</td>
<td>108</td>
<td>225</td>
</tr>
<tr>
<td>unique names</td>
<td>280</td>
<td>134</td>
</tr>
<tr>
<td>unique names which only responded</td>
<td>128</td>
<td>92</td>
</tr>
<tr>
<td>unique names which only originated</td>
<td>51</td>
<td>11</td>
</tr>
<tr>
<td>unique names which both orig &amp; resp</td>
<td>98</td>
<td>31</td>
</tr>
<tr>
<td>top 5 unique name volume</td>
<td>347</td>
<td>167</td>
</tr>
<tr>
<td>top 5 unique name original posts</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>top 1 volume</td>
<td>156</td>
<td>42</td>
</tr>
<tr>
<td>top 1 original posts</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>top 10 participant volume</td>
<td>504</td>
<td>261</td>
</tr>
<tr>
<td>top 10 participant original posts</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>unique names with &lt;=5 posts/name</td>
<td>225</td>
<td>107</td>
</tr>
<tr>
<td>total posts from unique names with &lt;=5 posts</td>
<td>489</td>
<td>201</td>
</tr>
<tr>
<td>total single posts from unique names</td>
<td>91</td>
<td>60</td>
</tr>
<tr>
<td>total single posts from unique names as originators</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>total single posts from unique names as responders</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>total single posts from unique names to total threads</td>
<td>91</td>
<td></td>
</tr>
</tbody>
</table>
Analyzing the texts

- From & Function on a spectrum
  - “Normative” to “Identity-oriented”
  - Differs from “discourse”, “style”, and “genre”
- Form
  - Modes of expression
    - Genre / generic structures (e.g., “article”)
    - Grammatical forms (e.g., “interrogative” or “third person”)
    - Literary forms (e.g., “irony” or “sarcasm”)
    - Cultural references (e.g., *Hitchhiker’s Guide to the Galaxy*)
- Function
  - Relevance to the community’s normative goals for the discourse event
  - Re-emphasizes that differences are qualitative
Visualizing Discourse Events

Identity-oriented

Normative

Function

Identity-oriented

Form
Newsgroups: comp.lang.lisp
From: namekuseijin <namekuseijin@gmail.com>
Date: Fri, 1 Aug 2008 09:23:07 -0700 (PDT)
Local: Fri, Aug 1 2008 11:23 am
Subject: Re: The Fundamental Confusion of Xah

On 1 ago, 07:13, Jon Harrop <j...@ffconsultancy.com> wrote:

> What proportion of your code base is written in Java?

Hey, this is Java we're talking about! Even interfacing to it probably demands a lot of glue code...

from: LSP0027.109
Hunting of the Snark

Google Is Your Friend

All Smart People Use Google

It Appears That You Are Not One Of Them
Self-regulation by public ridicule

plonk: excl., vt.

[Usenet: possibly influenced by British slang ‘plonk’ for cheap booze, or ‘plonker’ for someone behaving stupidly (latter is lit. equivalent to Yiddish, schmuck)] The sound a newbie makes as he falls to the bottom of a kill file. While it originated in the newsgroup talk.bizarre, this term (usually written “*plonk*”) is now (1994) widespread on Usenet as a form of public ridicule.
A complete picture

- Understanding technical communities requires synthesis of three frameworks for community analysis
  - Programming newsgroups...
  - ...Are fundamentally about discourse (Discourse Community)
    - Traditionally, as a text-based community
    - Topically, since programming is a linguistic task
  - ...Create knowledge (Epistemic Culture)
    - Programming methods
    - History of programming
    - What counts as programming culture
  - ...Evaluate, establish, and maintain practices (Community of Practice)
    - Programming
    - Community interactions
      - For both newsgroups and physical-space