Cassell, J., and Tversky, D. (2005). The language of online intercultural community formation. *Journal of Computer-Mediated Communication, 10*(2), article 2. http://jcmc.indiana.edu/vol10/issue2/cassell.html

The Language of Online Intercultural Community Formation

Justine Cassell
Technology and Social Behavior, Northwestern University

Dona Tversky
Media Laboratory, MIT

Go to a section in the article:

Abstract

This article examines how linguistic interaction patterns changed over time among a geographically and ethnically diverse group of young people in an online virtual community, the Junior Summit '98 online youth forum. The tools of word frequency and content analyses are paired with evidence from post-hoc interviews. Results demonstrate the ways in which these young people from different cultural, linguistic, and socio-economic backgrounds increasingly constituted themselves as a community, speaking in the collective voice, converging on a linguistic style, and concurring on the topics of conversation, the goals of the group, and strategies for achieving them.

Introduction

Almost paradoxically, technologies that allow people to communicate across great distances have allowed social scientists to make advances in understanding the construction and maintenance of community. In particular, information and communication technologies (ICTs) have provided a seemingly-miraculous window into the processes of community formation, including when the community members differ from one another along the axes of age, culture, economic benefits, language, and other dimensions that would hinder if not prohibit communication in the physical world. But how do online groups exhibit the hallmarks of community? How does their online language index patterns of assimilation to other community members, identification with group goals, growing engagement in joint activity? When a group of people from different countries comes together online at the same time in a new community grouping, are sub-groups distinguishable by their language use? Can one group be identified as the leader of linguistic trends? Are patterns of dominance among sub-groups in line with dominance in the physical world? How do such behaviors change over time as the members of the community come to know one another?

We approach these questions through an investigation of the "Junior Summit '98," an international virtual forum that brought 3062 children from 139 countries online to discuss global issues. The participants, speaking many different languages and representing a wide variety of economic and cultural backgrounds, discussed and planned ways to make the world better using technology. In order to analyze the tens of thousands of messages posted to the forum, we employ a set of research tools adapted from psychology

and sociolinguistics, including word frequency counts, content analysis, and in-depth interviewing, and apply them to this online context. In this article we discuss some of our results concerning language as an index of integration vs. the maintenance of separate cultural, age, and gender identities, by looking at how children from different backgrounds presented themselves online.

Background

The Junior Summit online forum was implemented to connect and empower motivated youth from around the world to make their voices heard on issues concerning young people. The organizers' initial goal was both activist and academic: to push the limits of the Internet to connect across distance (both social and geographic), to explore a technical design philosophy for the developing world of "designing to the lowest common denominator," and to research the effect of online community on "voice" (Cassell, 2002). In September 1998, when the summit was launched, it was the most extensive forum of its kind. Eighty thousand calls for participation, translated into 16 languages, were sent out worldwide to ministries of education, non-governmental organizations, and schools, with the goal of attracting participants with a passion for changing the world. The instructions read: "If you will be between 10 and 16 years old …we want to know how you see the state of children in your community and in the world, what changes you think can and should be brought about, and how these changes could be affected by the growth of the Internet and other new communication technologies." Any format for entry was accepted, including a video or photographic essay, a musical piece, a drawing or painting, or an essay in their native language. There was no requirement that the children have used a computer or the Internet previously.

Ultimately, the hosting institution, MIT, received over 8000 applications in 30 languages. With the help of international graduate students and faculty from across the campus, 1044 entries were chosen, based on a combination of geographic representation and the passion of the applicant for improving the state of children in the world. Table 1 shows the countries with the greatest representation, both in terms of participants and messages those participants posted to the online forum.

Latin America & Caribbean	Argentina	21	910	**Europe**	Croatia	11	65
	Bolivia	9	96		France	27	613
	Brazil	38	1428		Greece	32	3313
	Colombia	23	752		Lithuania	9	146
	Costa Rica	17	824		Romania	11	311
	Honduras	10	3		Ukraine	6	396
	Jamaica	20	1078		United Kingdom	14	1213
	Mexico	15	1231				
	Uruguay	19	174	**South Asia**	Bangladesh	8	51
					India	19	2734
North America	Canada	36	2521		Indonesia	5	471
	United States	67	4165		Malaysia	6	440
					Nepal	8	284
Africa	Cameroon	10	--		Pakistan	20	642
	Kenya	10	345		Philippines	6	449
	Namibia	9	28		Thailand	15	152
	Senegal	12	55				
	South Africa	30	1242	**Pacific Islands**	Australia	22	2416
	Uganda	8	55		New Zealand	17	2704
	Zimbabwe	9	20				
				East Asia	China	58	1746
Middle East	Israel	10	420		Hong Kong	8	242
	Lebanon	23	902		Singapore	14	1190
	Morocco	2	828		South Korea	8	206
	United Arab Emirates	17	2277		Taiwan	13	306

Table 1. Countries, children, and posts

Neither gender nor age was taken into account in judging entries. Nevertheless, the forum was divided almost equally between boys and girls (55% of participants were female), and the ages of participants ranged in a curve from 10 to 16, with the majority of participants aged between 14 and 16 years old.

Age	% by population	% of messages
10 years	2.7	1.2
11 years	6.7	7.2
12 years	12.4	10.2
13 years	11.6	17.7
14 years	23.8	23.8
15 years	23.8	28.3
16 years	16.7	7.4

Table 2. Representation and message contribution by age

Once the 1044 winning entries were chosen, participants were contacted with instructions and a CD containing software to allow them to participate in the Junior Summit forum. In addition, 200 computers were distributed to schools or community centers around the world for the use of participants, and 500 Internet subscriptions were also given out. When neither of these options was appropriate, the young people were reimbursed for using web cafés. The forum was implemented as a simple mailing list with the option of participating either by e-mail or through a web interface. Server-side were five translation engines to translate messages into either English, Spanish, French, Portuguese, or Chinese, as specified by each participant. These translation engines were modified versions of off-the-shelf software of the translation quality available in 1998, meaning that the output resembled gisting more than actual translation. To improve children's access to messages written in other languages, the participants were invited to translate messages for one another.

Important to the design of the summit was the philosophy that all participants be given equal access to the features of the forum, regardless of whether they were using a Pentium 4, or an Apple IIC. (For more details about the design of the forum, see Cassell, 2002.) For example, even after chat was implemented after several months, a system was put in place to relay chat messages to students with only e-mail access.

The main activities of the Junior Summit took place over a three-month period between September and December 1998. When the young people first logged on to the Junior Summit, they found themselves in homerooms, divided by geographic region. After four weeks, the participants suggested and voted on 20 topics to address, and divided themselves into those topic groups. After spending another two weeks in the topic groups, the participants elected two delegates per topic group to attend an in-person summit in the United States, in Boston. An additional six weeks were spent in the topic groups, preparing for the summit, after which point the 100 delegates left to spend one week in Boston. At no point was there more than one adult participating in each online group, and those adult moderators were trained to keep their participation to the absolute minimum-dealing with technical issues, and questions about the structure of the program.

Some of the participants dropped out when they discovered that they had not been elected as delegates, and some dropped out after the in-person summit. (Rates of attrition are presented in the results

section.) Many stayed on for an additional nine months, however, and some are still participating in the Junior Summit online community-for example, writing an online newspaper that has survived for six years. Importantly for the goals of this paper, the Junior Summit was a closed group of people-only those elected to the forum could access it-and the goals and structure of the forum were made explicit early on. Much like the imagined communities of nationalism described by Anderson (1991), these young people were told to think of themselves as a community, despite the fact that they had never seen one another. However, adherence to structure and participation in the stated goals were not policed by adults. Thus, the Junior Summit provides a particularly good opportunity for asking how the participants themselves constructed themselves-or not-as a community through their communication with one another.

Previous Literature

A growing body of research claims that it is no longer useful to think of community as only physical groupings of people. Rather, given the modern world in which we live, where easy and rapid transportation as well as telephone lines and email can sustain relationships, it is more appropriate to think of a community as a network of interpersonal ties that, like the isolated neighborhood communities that existed previously, provide sociability, support, information, a sense of belonging and social identity (Rheingold, 1994; Wellman, 2001; Wellman, Boase, & Chen, 2002). Studies of online communities have claimed that members exhibit behaviors that traditionally identify the presence of community offline. They make rules for themselves, elect leaders, and sanction misbehaviors. They identify as members of the group, and restrict membership to those who do belong. A number of researchers have attempted to set minimal criteria for online groups constituting virtual community (e.g., Herring, 2004). But what are the dynamics of the process? How does language function to construct a community where face-to-face cues and the exchange of physical goods are unavailable? How does language index changes in the nature of a community over time?

The literature is particularly sparse on the topic of how language functions in cross-cultural communities online. Previous researchers have investigated how different alphabets are adapted for use in ICTs; for instance the way in which Arabic is transformed into Roman characters in instant messaging (Palfreyman & Khalil, 2003). The use of different languages on the Internet has also attracted attention, with researchers exploring how and why English continues to be the lingua franca of the Internet (Durham, 2003) as well as when machine translation can help support other languages and when it cannot (Climent, Moré, Oliver, Salvatierra, Sànchez, Taulé, & Vallmanya, 2003). Most of the studies that have looked at cross-cultural interactions using ICTs have been limited to two cultures and self-reported data. For example, Ma (1996) tested five propositions about culture and computer-mediated communication (CMC) through written reports and interviews with American and Asian university students who had participated in a relay chat system. Based on the self-report data, Ma argues that both East Asians and Americans were more direct in CMC than in face-to-face communication. Interestingly, he reports a discrepancy in self perceptions among the groups-even though East Asians thought of themselves as direct online, Americans still found them polite and reserved. Similar methods were employed by Meagher and Castanos (1996) in an attempt to look at how perceptions of Mexican and American culture are modified by participation in online exchanges. The authors measured changes in attitude toward one's own culture and the other culture using a questionnaire before and after a cross-cultural exchange.

Transcripts were also referred to for supporting evidence, although no results from systematic coding were reported.

Research on language use in the physical world has shown that language use does change over time, and does differ across communities. Thus, as people get older, they use more positive emotion words, fewer negative emotion words, fewer first person singular self-references, more future tense, and fewer past tense verbs (Pennebaker, Mehl, & Niederhoffer, 2003). People also show more cognitive complexity in their words as they age, using more causation (because, effect) and insight words (think, know, consider). Differences in the way women and men present themselves verbally have been the subject of much debate, and little consensus. Lakoff (1975) argued that women speak in a less assertive manner evidenced by a greater degree of politeness, more hedges and intensifiers, and more "tag questions." In contrast, men are purported to be briefer, more direct, and less emotional in their choice of words. More recently, Eckert and McConnell-Ginet (1992) have problematized the issue of gender and language by looked at gendered ways of speaking as arising from intersections of class, race, gender, and particular discursive contexts.

The standard for analyses of language use online comes from Herring (1996), who looked at one female-dominated and one male-dominated listserv. She found that both men and women structure their messages so the exchange of views is paramount to imparting information. Although messages posted by women are more interactive, they also contain more information while men tend to express their views (often critically) more often. Herring's research further suggests that the minority gender on each listserv conforms to the style of the majority. Panyametheekul and Herring (2003) extended this research to look at the interaction between gender and cultural origin. In their analysis of turn-taking in a Thai chat room they found patterns of interaction unlike those found in previous literature, which focused on predominantly Western online communities. Thai females both participated more often than males and received more feedback to their messages, contradicting common expectations about the roles of women in Thai society, and the role of women online. Herring's work is often cited as evidence for the "re-construction" of physical categories such as gender in the apparently disembodied space of the Internet.

Much research on online communities has been aimed at educational outcomes. Particularly relevant work has been done by Rourke, Anderson, Garrison, and Archer (1999) on "social presence," in their view a necessary feature of a successful online learning community. Rourke, Anderson, and Archer (2001) measure social presence using three types of communicative responses: interactive responses, affective responses, and cohesive responses. To date the research has focused primarily on development and testing of the coding tool. Jones (1997) presents an archeological approach to studying online communities, arguing that the cultural artifacts and physical traces created by a community can be examined to see if they fit the requirements of a true community, for which he sets forth four conditions: (1) a virtual common-public-space where a significant portion of interactive group-CMC occurs; (2) a variety of communicators; (3) a minimum level of sustained stable membership; and (4) a minimum level of interactivity. Another methodological approach comes from Herring (2004), who outlines an approach for the empirical analysis of on-line verbal interaction. She places an emphasis on the allied use of qualitative and quantitative language-focused content analysis, and describes a range of techniques

that can be adapted to different research questions. Content analysis has been employed since the 1950s, and was extensively employed by Flanders (1970) and Sinclair and Coulthard (1975) to study patterns of verbal interactions between teachers and students in the classroom. One problem reported by researchers using these observational tools is the labor required to transcribe real-time, face-to-face interactions, the labor required to develop bottom-up (non-a-priori) coding schema, and the time and energy required to code the transcripts.

Although debates about children on the Internet rage in the popular media, few studies have shown conclusive findings about the relationship between children's use of ICT's and social and psychological outcomes. The famous Internet Paradox study (Kraut, Patterson, Lundmark, Kiesler, Mukopadhyay, & Scherlis, 1998) initially demonstrated adverse psychological effects of Internet use-increased loneliness and depression with increased Internet use-but these effects disappeared with time. Reviews of the literature on social outcomes and ICT use among children show few documented social effects, either positive or negative (Kraut et al., 1998; Kraut, Kiesler, Boneva, Cummings, Helgeson, & Crawford, 2002; Subrahmanyam, Kraut, Greenfield, & Gross, 2000), leaving both a lacuna and a serious need for studies that address the effects of participating in online communities such as the Junior Summit.

In the present work, we adopt a perspective that understands the relationship of participants and their online community to be mutually constitutive, and highly context-dependent. That is, rather than asking *whether* the Junior Summit adheres to an external definition of a "true community," we ask what is the *nature* of the particular community that the Junior Summit participants have constructed at any given time, as indexed by their communicative behavior online. The organizers of the Junior Summit decided on the aggregate of people who were to be participants, and set them a common enterprise. As researchers, we can then ask if the young people developed a shared way of talking and doing things, engagement in shared values, and what those practices tell us about the participants' conception of community.

Methods

The data sets that comprise the Junior Summit are of three types: (1) the 48,000 messages posted to the online forum for the period September 1998-September 2003; (2) in-depth interviews about the effects of the Junior Summit conducted with 78 participants from 20 countries five years after the Summit began; (3) questionnaires on socio-psychological variables (primarily self-efficacy, meaningful instrumental activity, social networks) filled out by the same subset of 78 of the children five years after the summit began.

In this article, we discuss results from analyses carried out on a subset of this large data set: (1) all of the messages posted by those children who participated independently (as opposed to as a part of a team or group of children) and who chose English as their primary language of communication over a three month period, and (2) in-depth interviews with 37 children in 17 countries. We employ two types of analyses to interpret the email messages-a computational word frequency analysis, and a more sensitive content analysis carried out by human coders. In the following sections we describe our methods in greater detail, and explain the selection of the data.

Word Count Use over Time

As discussed by Pennebaker, Francis, and Booth (2001), word frequency can be a powerful tool in understanding the psychological and sociological profiles of individuals and groups. Publicly available software based on Pennebaker et al.'s Linguistic Inquiry Word Count technique analyzes text files on a word-by-word basis by comparing each word in a given file to words in an internal dictionary representing a variety of different psychological or linguistic dimensions. The categories include basic linguistic measures such as frequency of nouns, pronouns, and articles as well as more complex ones that tap into psychological processes, such as linguistic measures of anxiety and sadness. In this way, LIWC provides a fairly straightforward index of how individuals use language, in a way that may serve as a marker of a wide variety of individual differences and commonalities, including demographics and personality.

The data set we submitted to LIWC analysis includes all the messages posted in the first three months of the forum by individual participants who *chose* to write in English, amounting to a total of 19,004 messages. These messages were written by 374 participants, 200 of whom were girls and 174 of whom were boys. Girls wrote 62% of the total messages while boys accounted for 38%. Although all the subjects in this sample chose to write in English (for some it was a second or third language), the sample includes children from 94 countries and every region of the world, with ages ranging from 10 to 16. For this particular article we chose to analyze only messages written in English because we do not yet have the resources to analyze word frequencies in the other languages, and translation would add confounding factors since the words counted are not those originally chosen by the writers. Future word frequency analyses of forum messages will look at non-English posts as well.

Our analysis employed most, although not all, of the LIWC categories, and we also added several of our own, including hedges, WH questions ('who,' 'where,' 'when,' etc.), apologies, and several others. These categories were added because they are examples of language that does social work, as opposed to being purely informational. Similarly, examples of LIWC categories that we will discuss here include first person singular and plural pronouns, negations, assent, positive emotion, and reference to the future tense. In order to look at changes in the linguistic features over time, each feature was standardized by time period and participant. For example, one of the features in the LIWC set is "first person plural pronouns" which includes 'we,' 'our,' 'us,' etc. For the "we" feature we created a score for each participant by taking the total number of "we" words used in each period and dividing it by the total number of words written in each period for that user. This gave us a feature density for each participant in each time period. Next, we calculated the mean of this new "density" variable, and normalized it to one by dividing it by its mean so that the mean for each feature is always one. We refer to this value as a "normalized feature score" (NFS). The NFS allows us to determine quickly if a time period or a sub-sample of the population is above or below mean, and to track changes over time. Thus, for example, if children from Southeast Asia have a "we" score of 1.07 NFS, we know that they speak in the collective voice more than the general population; we then look at standard deviations and other aspects of the data to determine if that difference is significant.

A More Content-Oriented Approach

The LIWC is capable of capturing many aspects of an individual's writing style, but only those that can

be explored through the frequency of particular lexical items or groups of words. Thus, in addition to LIWC analyses, we used a methodology that allowed us to concentrate on the content of the participants' messages. For example, categories such as "giving feedback on an idea" cannot be captured through analyses of single words, but are an important index of involvement with others. Our content analyses therefore addressed questions such as how the children proposed new ideas, whether they gave feedback to one another, and what was the nature of their feedback.

No previous work captured the granularity we hoped to achieve with our analysis, and thus after looking at work by Bales (1951), Herring (1996), Rafaeli and Sudweeks (1997), and Rourke et al. (2001), we ultimately developed our own codebook. In addition, because we hoped to capture the ways in which the participants themselves chose to constitute community through language, we did not start off with an *a priori* list of content categories. Instead, using a Grounded Theory-inspired methodology (Strauss & Corbin, 1994), in which codes are inductively and iteratively derived from the study of the phenomenon represented, we developed a 34-feature codebook to capture the ways in which participants express ideas, give feedback to peers, and present themselves online. Codes were applied to an entire message, and more than one code could be assigned to a single message.

The codes fall into two general categories: (1) "informative"—meaning that the utterance conveys information, and is able to stand on its own, as in the case of an idea or an opinion; and (2) "interactive" or "interpersonal"—meaning that the utterance is in some way a response to the contribution of another writer (Rafaeli & Sudweeks, 1997).

Informative: present solution, present extensive solution, provide global knowledge, express strong opinion, express opinion with hedging, share personal narrative, share biographical information, synthesize

Interactive or **interpersonal**: agree, agree add ideas, disagree, counterpoint, acknowledge input, ask for information, invite feedback about an idea, thank, offer advice, sympathize, humor, express disappointment, delegate work to the group, rally cry, negative rally cry, volunteer

Listed below are three examples of the categories in the codebook we developed, along with their definitions and examples. The first two are examples of *informative* codes, while the third is an example of an *interactive* category.

- **PNAR**—Share personal narrative
  Tell a story or provide personal experience, knowledge directly related to the topic
  Ex. "I know quite a few people who say education is boring and expensive," "We have to sing a song called 'KIMIGAYO,' a kind of national anthem."

- **BIO**—Share biographical information
  Offer personal information not related to task about themselves (1-3 sentences)
  Ex. "I like to fish." "I love to listen to Metallica." "As you may know I have been fasting."

- **AG+**—Agree and add ideas
  Agree with or praise someone's ideas and add one's own idea(s) to that of peer
  Ex. "I agree about your plan. And I also think we could raise money by …"

Since the content coding was done by hand, and was therefore far more time-intensive, this round of coding was carried out on a subset of the data: the complete set of messages posted by 36 users who represent a spread of geographical representation and level of participation in the program (as measured by how many messages they posted). These participants came from 15 different countries and include 23 girls and 13 boys. The total number of messages they produced, which we then analyzed for content themes, totals 4377.

It is clear that an analysis of 36 children's messages will not allow us to examine regional differences, but it does allow us to look at change in patterns of participation over time. Because the first three months of the Junior Summit asked children to regroup several times (in homerooms, in topic groups, in delegates), and in order to look at the effects of these particular events on communication patterns, for the purpose of our analyses we therefore divided the summit into time periods, as shown in Table 3.

Phases of the Forum	Duration	Time Period
Homeroom	4 weeks	T1-T2
Topic Groups	8 weeks	T3-T8
Elections	1 week	T3/T4
In-person conference in Boston	1 week	T6

Table 3. Phases of the forum

Coding Reliability

Inter-rater reliability on content coding was assessed for the team of three coders, and Cohen's kappa scores were calculated for each code. The kappa score of the individual codes (for example 'biographic information' or 'personal narrative') ranged from 0.22 to 1.0 with a mean and mode of 0.66. Codes that had low inter-rater reliability, or were very infrequent, were omitted from the analysis. Only five out of the 32 codes had kappa scores below 0.5; none of those are discussed here.

In this article, we report on content analysis features that appear in a minimum of 5% of messages. This choice excluded a handful of codes from our codebook, some of which are interesting for their absence. For example, the codes 'give advice to peers,' 'express sympathy,' and 'synthesize discussion' were too rare to report reliably. In addition, and perhaps surprisingly, none of the codes capturing antagonistic or even mildly critical communicative actions-criticism of ideas, disagreement with peers and expressions of disappointments with the program-occurred with sufficient frequency to report. While flaming is common in online fora, it appears that this forum was unique in that participants either refrained from or were uninterested in voicing negative feelings toward one another or toward the program as a whole. There are any number of reasons for this, not the least of which is that the participants may have been 'campaigning' at some level for the conference in Boston, though it should be noted that these features appear with the same infrequency following the election announcements.

Coding Interviews

Five years after the start of the online forum, extensive follow-up interviews were conducted with a sample of 78 of the original participants. The subjects for the follow-up were selected based on three criteria: first, a minimum level of participation in the forum as measured by the number of messages they posted; second, geographic diversity; and finally, other diversity issues such as gender, rural vs. urban, older participants vs. younger. While it originally seemed worthwhile to interview children who did not log on often to understand what deterred them from participating, we soon found that low-participation children hardly remembered the program at all, and their few memories were not detailed enough to be informative. In practice, this meant that we interviewed young people who had posted a minimum of 50 messages and up to 1000 messages over a three-month period.

The interviews relied on a set of open-ended questions, and lasted between two and four hours. The first half of the interview concerned general questions about summit participation and effects, while the second half of the interview began with giving the young people their entire set of posts (that is, all of the emails sent by the child to anybody from the beginning of the forum until the child stopped participating) and then asking the interviewee to reflect on what was going on as s/he participated. So as not to put words in the participants' mouths, the interview protocol did not contain any direct questions about community formation or intercultural communication, but these ideas emerged when interviewees talked about how their attitudes had changed because of the Junior Summit and the benefits they felt they received. While the majority of interviews were conducted in English (by the interviewee's request), an interpreter was always offered, and was accepted on several occasions (Argentina, Bangladesh). The questions were designed to elicit goals for participating in the Junior Summit, the context of the child's participation at home and at school (positive and negative feedback about participation from family, peers, school, assistance in participation, effects of participating), both positive and negative evaluations of the program, to gauge impact on later life choices, and to assess effects on social networks. After interviewees were shown all of their e-mails to the Summit from five years earlier, they were asked to reflect on the structure of the online community, their interactions with other participants, the voting process, and the reasons they stopped participating.

Audiotapes from the interviews were transcribed and coded for common themes, using the same bottom-up method for deriving categories to code. Ultimately, a codebook containing 245 codes, capturing participants' impressions of and experiences with the Junior Summit, was created using the children's own words. Because coding interviews that ran between 40 and 170 pages in transcribed length is time-consuming, for the current article we analyzed data from 37 interviews (24 girls and 13 boys) in 17 countries, selected at random.

Results

In what follows we first look at a number of potential indices of group convergence, using both the word frequency and content analysis approaches to our data. After discussing convergence, we then turn to divergence: differences by region, age, and gender, as well as questions of mutual influence. We then turn to our interviews with the participants five years after the forum to discover their own perspective

on how and whether the Junior Summit became a community, and what that community meant to them. Finally, we address some of the ramifications of the influence of particular cultures on others, before concluding with remarks on the general lessons to be drawn from our results.

Cross-Cultural Community Constitution

As described above, several studies have examined virtual communities online, but none have looked at a group of people as diverse as the Junior Summit. Our first analyses therefore concern the most basic indices of whether this group of people considers themselves to have any commonality at all-to constitute a group. In this context we look at the most basic marker of group vs. individual, the use of "we" vs. "I." The choice of pronouns gives us one very simple index of identification with others; a more complex and ultimately more interesting index comes from the choice of topics over the course of the first three months of the Junior Summit. Did the participants converge on a finite set of topics, and ways of speaking with one another, or did their conversations range over all and every subject possible? Did the range of communication styles change over the first three months? We end this section by looking at the children's own perceptions of the value of community membership.

Personal Pronoun Use as a Signifier of Community

To see if the participants identified themselves as a group, we first examined the use of personal pronouns over time, hypothesizing that if the Junior Summit felt like merely a gathering of individuals and not a true community, there would be no change in participants' use of individual and collective pronouns throughout the three-month period.

In effect, as time wore on over the first 12 weeks of the forum, children decreased references to themselves as individuals, and increasingly spoke with a collective voice—they used "I" less and "we" more. More specifically, "we" words (lets, let's, our, ours, ourselves, us, we, we'd, we'll, we're, we've) increased by 16% from period 1 (0.81 NFS) to period 2 (0.94 NFS)—that is from the first two weeks to the second two, and by another 7% over the following two weeks to period 3 (1.04 NFS). The "we" feature reached a final peak in the 6th time period at 1.25 NFS. The data points and regression line are plotted in Figure 1. (We: coef= 0.0425; std. error= .0054; t= 7.79; $p<0.000$ [CI] = -0.0318-0.0531.)

Conversely, a regression analysis of the use of "I" words (I, I'd, I'll, I'm, I've, me, mine, my, myself), showed a significant decrease over time (I: coef = -0.0159; std. error= .0020; t= -7.89; $p<0.000$ [CI] = -0.0199-0.0120). By looking more closely at the data points plotted in the graph, we find that there was a significant decrease by 4% between the first (1.07 NFS) and second time periods (1.03 NFS), and then again by 5% between the second (1.03 NFS) and third time periods (0.98). First person singular pronouns hit their nadir in the 6th period at 0.9 NFS. (See Figure 2.)

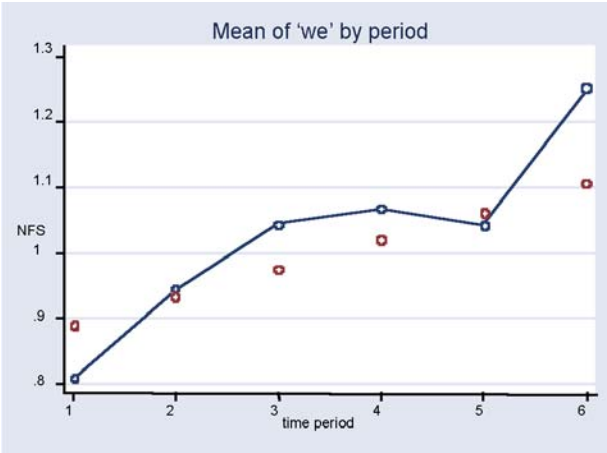


Figure 1. Collective first person pronouns

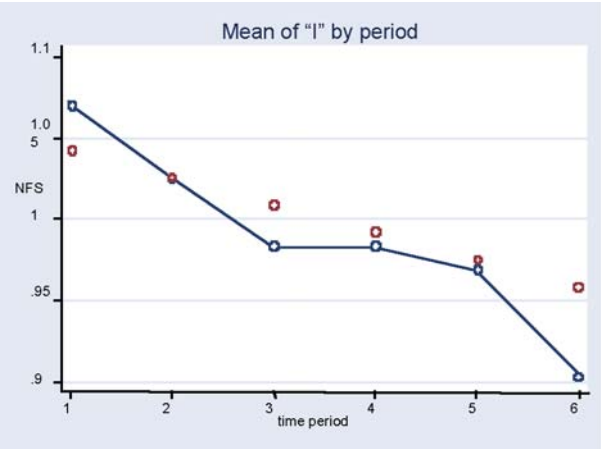


Figure 2. Singular first person pronouns

In addition to demonstrating individual versus group identity ("I" vs. "we"), pronouns are thought to be indicative of people's level of focus or involvement with others (Pennebaker et al., 2003). It has for example been found that people decrease their use of "I" and increase their use of "we" in periods of shared trauma (Pennebaker & Stone, 2003). In one study of personal pronouns in an online chat room before and following the death of Princess Diana, researchers found that after Diana's death, the use of 1st person plural increased by 135% and the use of "I" dropped by 12%. The effect lasted ten days before pronoun use returned to normal.

If children are increasing references to themselves as a member of a group and decreasing references to themselves as individuals over time, is this evidence of community-constitutive behavior? Possibly, but not without further investigation. Considerable attrition took place over the course of the online forum, and the community-constitutive effect suggested by the "I" and "we" word trends may in fact be due to participation trends. For instance, it is possible that children who spoke about themselves using "I" words dropped out over the course of time, leaving those who used more "we" words to stay in, thus increasing the relative incidence of "we" words in the forum.

We examined this question in several ways. First, we looked to see if those children who dropped out used more "I" words than their peers who stayed online, and we did find this to be true. The ratio of "I" words to "I" + "we" words was 89.5% for children who dropped out and it was 87.4% for those who stayed online ($F(1) = 21.52$ $p < 0.05$). However, when we ran the regression analyses again using only the participants who stayed online for the whole initial three month period, we found that the community referencing effect-in which "I" words decrease with time and "we" words increase with time-still held true. That is, the new regression coefficient for "I" still showed a significant decrease over time (I: coef= -0.0234; std. error= .0003; t= -72.34; p<0.000 [CI] = -0.0240- -0.0228) and the regression coefficient for "we" still demonstrated significantly increase over time (we: coef= 0.0724; std. error= .0008; t= 81.27; p<0.000 [CI] = 0.0707-0.0742). In fact, the contributing effect of children's attrition from the forum comprises about 18% of the total effect for "I" pronouns and about 1.7% for "we" words.

In order to ensure that the effect truly held for the individual participants of the forum, we ran one additional linear regression in which we defined the proportion of "I" pronouns to the pronouns "we" and "I," for each child for each time period, and found that the ratio of I to I + we, again, decreased significantly with time (I: coef= -0.0060; std. error= .0006; t= -9.84; p<0.000 [CI] = -0.00717--0.0048).

It is clear, then, that children who remained as participants of the Junior Summit did decrease reference to themselves as individuals in favor of references to a group. However, is the group referred to the Junior Summit, or do children use "we" to refer to their families, their peers at school, or their countries? In order to examine this question, we took a random sampling of messages from each of the six time periods and checked the referents of "we." From this sampling, it appears that first-person plural pronouns occur most often in one of two instances: either children use it to describe a familial or cultural custom, e.g., "Sorry I haven't written in a while, we are having exams here and I don't have much time…" or in reference to the Junior Summit community, e.g., "If in schools we can teach the kids to share, to care for each other, to live as one, and not as separate individuals with their own demands uncaring for each other, then we can make things better." Over the course of the summit, the use of "we" shifts from referring to the child's local community to the global online group.

The trend is observable within the following messages of one writer over time. Instances of "we" in reference to the writer's home community decrease as soon as the group starts planning for action. The subsequent instances of "we" then sometimes refer to the entire Junior Summit community and sometimes to a smaller subset working on one issue. (Note: spelling and typography are preserved from the original messages.)

> Sept. 4
> HI! I'm MC from Malaysia…Most Malaysians are bilingual or trilingual much like our Singaporean neighbours. Some of us can even speak 4 or 5 languages and dialects. …I look forward to swapping ideas with all of you during the forum and making many new friends. It would be a dream come true if we could help to solve one of the World's many problem.
>
> Sept. 7
> Furthermore, Malaysia is currently facing the Economic Crisis which we are trying hard to fight. …About the economic crisis, although we might not be able to do much, we could perhaps come up with a few suggestions to solve the problem. If the world's leaders were to listen to these, it just might work. As participants of the JRSummit, we should be more optimistic. …
>
> Sept. 14
> I sincerely hope that with more ideas contributed by my fellow members of home 24 we can make the world look more seriously into the problems of illegal immigration and hopefully our world will be a better place for all to live in, irrespective of race, colour, or creed.
>
> Oct. 4
> Dear J, I read your e-mail after sending off my latest e-mail. I am astonished to find that we have many ideas on the uniformity of cultures in common. Great minds think alike, eh? :-)
>
> Oct. 13

Thanks to both of you for leading me to the Esperanto and International Phonetic Association websites. Both languages have taken more than 100 years to reach its present status. If we were to try to create another, we won't live long enough to see its success! It's sad that politics and nationalism (fanaticism?) have been restricting their development.

Oct. 22
I agree with you that we have to save cultures other than our own. After all, it is only fair that we help cultures that are in as much danger of losing their identity as ours. That is why we are here in the Junior Summit. What we need now are more ideas no matter how crazy.

As another index of the children's involvement with the others and with the group, we looked at WH questions (who, where, what . . .), and talk about the future. Both show increases over the course of the Junior Summit forum from the homeroom stage to the point where 100 of the participants leave for Cambridge (WH questions - coef = 0.0494; std. error= .010; t= 4.98; p<0.000 [CI] = 0.0300-0.0689; Future coef = 0.0341; std. error= .0060; t= 5.60; p<0.000 [CI] = 0.0021-0.0460). This result, paired with the previous one, seems to indicate that over time the children spent less time engaged in description of themselves, and more time engaged in showing interest in one another. In order to examine whether this was indeed the case, we turn to analysis of the content of the children's messages.

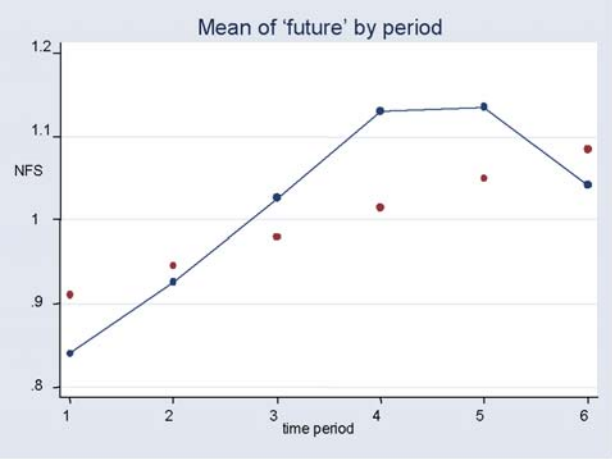


Figure 3. Future words

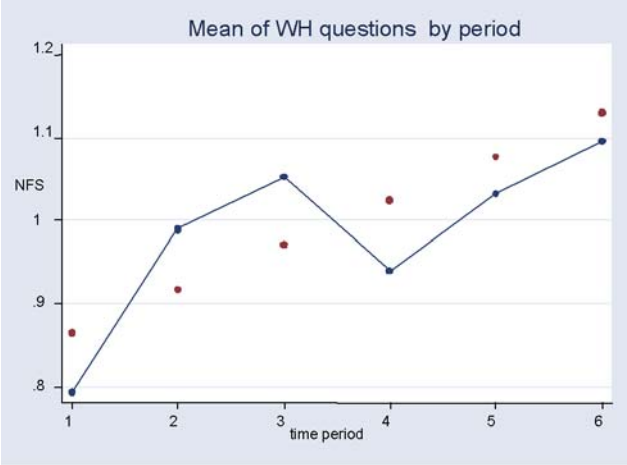


Figure 4. Wh questions

Topics of Conversation as a Signifier of Community

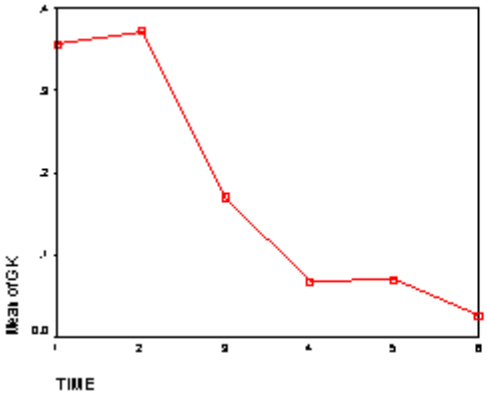
Whereas the previous analysis relied on word frequencies to show increasing references to the community over references to the self, below we use our analysis of the content of the children's posts to look at *how* these discussions took shape. We look at each of the first three months in turn, and find that each can be characterized in terms of the nature of the exchange among the participants.

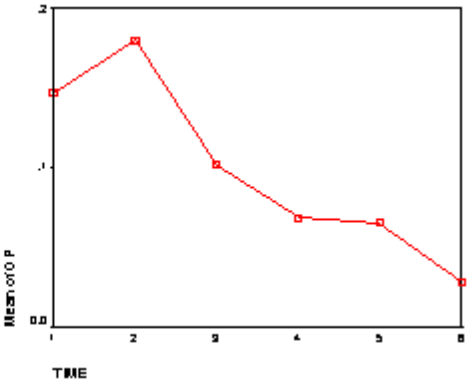
Month 1: Information exchange

Results of the content analysis show that in their first month online the children introduced themselves and started raising ideas for problems to work on during the forum. They voiced opinions about issues and supported them with stories from their own experiences (what we refer to as "personal narratives")

or information learned from outside sources ("global knowledge"). Most of the forms of communication that dominate this period are informative as opposed to interpersonal in that they do not directly respond to another person. Interestingly, all of the features of communication that peak in the first month, including opinion, solution, personal narrative, and global knowledge, then follow a similar pattern in showing significant decreases in frequency over the next five periods ($p<0.05$).

Below are graphs showing the frequency over time with which children gave ideas supported with global knowledge (GK) and gave opinions about topics (OP). As can be seen, during the first month children engaged in a veritable deluge of information sharing.





Figures 5 & 6. Global knowledge / Opinion

Although this period of time was dominated by information-based communication, there is evidence to suggest that the children were also learning how to converse with one other across cultural boundaries. For example, both straightforward objections to another person's idea ("I don't think that is a good way to go") as well as more diplomatically-stated differences of opinion in which children first praise an idea and then criticize it ("That is a nice thought, but wouldn't it be better if...") diminished after time 2 (Disagree: $F(4616,5)= 3.95$ $p<0.001$; Counterpoint: $F(4616,5)= 7.22$ $p<0.000$). Meanwhile, positive feedback began to rise, and was the most frequent sort of feedback after time 3 (Agree: $F(4616,5)= 7.26$ $p<0.000$).

In addition, and not surprisingly, during this early period of the summit, the children were also interacting and getting to know each other on a personal level. Thus, one other characteristic of the first two weeks of the forum was a discussion of children's personal lives. For example, when they first came online, children spent a message or two telling their fellow participants about where they came from and what they liked doing. The underlined section of the following message is an example:

> Namaste!Hi!I am Deepak from Bombay (India) . I study in Rajhans School. My hobbies are collecting old coins and making electronic projects . I feel strongly against the exploitation of children and feel great that I am given an opportunity to combat this menace. I would feel great if you would send a reply . Bye!

However, while it is commonly thought that online communities, particularly for young people, are primarily focused on this kind of small talk, i.e., chat about personal lives and other non-task related

common ground (politics, movies) in order to bond or become committed to a group, our data suggest otherwise. The amount of time spent discussing personal affairs started out high during the first week but immediately began dropping rapidly until the very last two weeks, when some of the participants were in Boston. Recent research suggests that social messages can interfere with work or learning in online communities (Rourke & Anderson, 2002), and the participants of the Junior Summit seemed intent on constructing a community around shared work rather than social relationships alone.

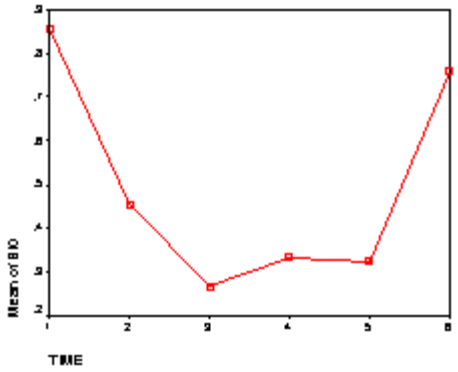


Figure 7. Biographic information

Content codes that peak in ...		
1st month	**2nd month**	**3rd month & conference**
Opinion	Agree (Positive feedback)	Apologies
Solution	Agree and add ideas	Action plan
Global knowledge	Thank	Delegate
Disagree	Acknowledge input	Volunteer
Personal narrative	Rally	Humor
Bio	Request feedback	Bio

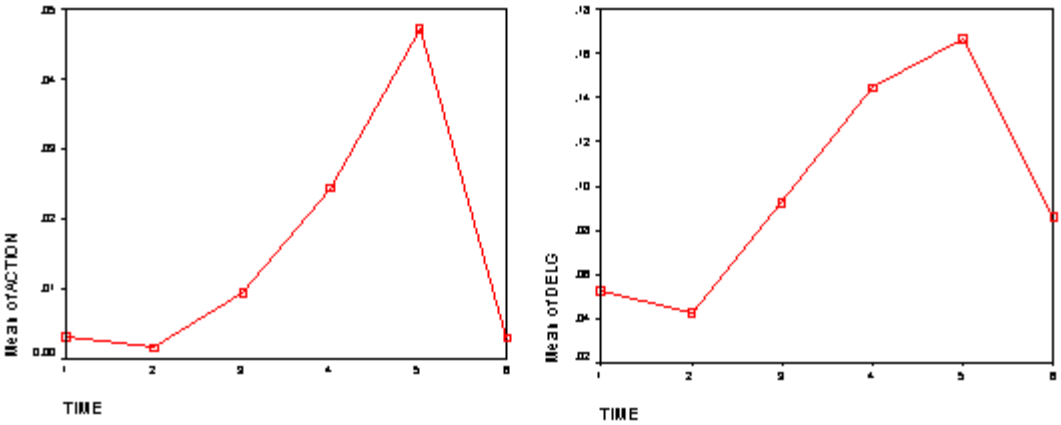
Table 4. Stages of the forum in terms of content

Month 2: Interaction

More messages were posted during the second of the initial three months online than during any other time, and this period also demonstrated the most *interactive* exchanges between participants. Whereas in the first month, children spent their time suggesting novel ideas or solutions, in the second they primarily modified each other's ideas, further developing each other's plans and working collectively. As noted above, negative feedback or disagreements appear rarely during this period, replaced instead with abundant positive feedback and appreciation for one another's opinion. The participants also requested feedback from one another frequently. In addition, rallying cries to the group were common during month two, signifying a desire for the group to act together, and a commitment to one another. Some of this positive energy may also have served as a way of garnering support by making apparent participants' dedication to the community-note that voting for delegates to travel to Boston took place at the end of week six and that delegates were announced at the beginning of week eight.

Month 3: Planning for the real world

After delegates to Boston were announced, a number of participants left the forum as their hopes of a trip to Boston were dashed. Many non-delegate participants remained, however, and one might think that it was the most dedicated young people who chose to stay. During this period, each topic group was tasked with creating an action plan, and as we see in Figure 8, discussion of the action plan and delegating work to one another went hand-in-hand during the third month. In this third month online, the children not only delegate to one another but also volunteer for tasks themselves-including offering to build a website for the group or write a wrap-up report on the discussions of the past week. This introduction of action into a forum that has primarily been about talk supports the interpretation that shared talk has led to shared goals, and now to shared practice in the world. The construction of the community has achieved the point where the assignment of roles-to one another and to oneself-has become possible. The increased use of rallying cries over time also suggests that by month three there was something around which to rally.



Figures 8 & 9. Action plan / Delegate

Finally, one other code that supports the notion of increased community feeling based on shared action is the use of apologies, which rose steadily from month one to month three (Apologies: F (4616,5=5.00; p<.000). Apologies usually took the form of "sorry I haven't written in 2 days." Note that apologies of this sort are one way of indexing one's participation in a project, as regret for one's absence is signaled.

The results from the frequency count and hand-coded email message analyses presented here suggest that the diverse participants construct themselves as a community via their use of language, and the way they work together. In this context, the first stage of marking themselves as a community seems to be speaking as a collective voice ("we" instead of "I"), the second stage is marked by interactive patterns of conversation (giving feedback or responding to each other's ideas), and the third stage is proposing action in the world based on an established commonality of goals.

Homogeneity or Diversity in the Community

The language behavior data, then, point to the Junior Summit being self-constructed as a community, in addition to being named a community by the organizers. Thus far, however, we have not investigated

how differences among the participants played out-differences in origin, for example-and whether those differences persevered over the course of the forum.

Regional differences

From Murdock (1945) to Hofstede (1980), there exist a number of different ways to carve up the world into groupings of putatively similar cultural practices. For each of these there exists a correlative theory of communicative behavior by cultural grouping (e.g., Triandis, 1989). In keeping with our practice of steering away from a priori coding schemes while still making a first pass at investigating differences among the young people of the Junior Summit, we grouped participants into regional divisions-supra-categories of the regional divisions that Junior Summit organizers used to group participants into homerooms during the first two weeks of the forum. These regions were Europe, the Middle East, Africa, East Asia, Southeast Asia, Central and South America and the Caribbean, North America, and the Pacific Islands. We then conducted a stepwise regression analysis on the word frequency counts reported above. The regression analysis included interaction terms between time period and region indicators so that each region had its own intercept and slope and we could compare each region to the others as time progressed. Europe was chosen as the reference group because it was the largest, though this choice has no effect on the differences reported for each region. The features reported below are: singular and collective pronouns ("we" and "I" words), apologies, references to the future, hedges, and WH questions (who, what, where, etc.).

- **Personal pronouns**: The population trend showing that "we" words decrease and "I" words increase with time holds true for each of the regions independently as well. That said, there are some differences in the frequency with which each region used the pronouns. A regression analysis showed that North American children used "we" words significantly less than the mean, and also increased their use at a rate slightly slower than their peers. Meanwhile, children from Central and South America started at a lower point but increased at the same rate (I: coef (Europe) = 0.0463; std. error= .0056; t= 8.26; p<0.000 [CI] = 0.0353-0.0573).

  There were also some differences in the way regions used or stopped using singular first person pronouns. Pacific Islanders and children from Central and South America and the Caribbean, as well as those from North America, started out using "I" significantly less than the mean and decreased in parallel with the general population. In contrast, Asia used "I" words significantly more than the mean, starting at a higher point and decreasing their usage at a faster rate than the others. (I: coef (Europe) = -0.0132; std. error= .0021; t= -6.21; p<0.000 [CI] = 0.0174-0.009)

- As described earlier, **Apologies** increased over time. Children from Asia, however, increased their use of apologies more rapidly than the others.

- **Future**: There were no regional differences in the amount children referred to the future although, as reported above, all groups increased use over time.

- **Hedges**: Middle Eastern participants used significant more hedges initially than did other children, while children from Africa and Central/South America/Caribbean used fewer initially than the others.

- As reported above, the use of **WH questions** increased with time. The Pacific Islands and Africa,

however, stayed at a higher level of WH question use than the others, while the Middle East used these terms less frequently. In addition, the Middle East and North America increased more rapidly in their WH questioning than the others.

While each of these results could form the topic of an entire research investigation into patterns of language use among young people from different countries, here we use this division as one way to look at divergence and convergence over groups, as opposed to over time. The results just described, then, serve as evidence that not all of the participants on the Junior Summit behaved in identical ways during the forum. Since change over time risks masking differences at each time, we first establish difference and then look at influence of the groups on one another.

Mutual Influence

People in conversation *entrain*, or match each other's words or linguistic styles (Niederhoffer & Pennebaker, 2002); a similar process also occurs in online discussions (Herring, 1996). In a global conversation of this size, how are the different populations-by culture, age, gender-in the forum influenced by one another? Do they converge in aspects of their language use? Do they diverge? Is one subpopulation dominant? Tracking behavior by region over time allows us to look at how the groups interrelate. That is, do the regional groupings uncover convergence, divergence, or independence of behavior over time? Convergence and divergence can be measured by subtracting the differences between regions at the end of the six periods from the differences observed at the start to determine if they were meaningful.

Convergence: Singular First Person Pronouns

The community as a whole showed important convergence in their use of "I" words over time. Children from the Pacific Islands, North America, and Central and South America and the Caribbean all started significantly below mean and decreased at a constant rate, while East Asia, which started out at higher use of "I" words, decreased faster than the other regions and converged with the remainder of the population.

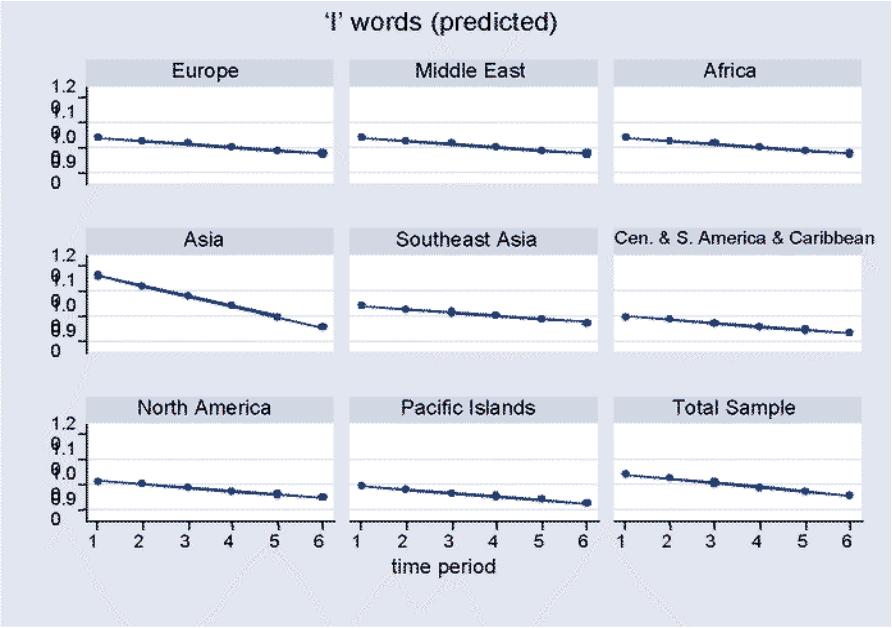


Figure 10. Singular first-person pronouns by region

Divergence: Apologies and Collective First-person Pronouns

Within the sample of features we have used as indices of community mindset, children showed diverging use of two-apologies and first person pronouns. As corroborated by the manual-coded data, the amount children apologized increased over time. Within the group, though, children from East Asia increased at a faster rate than the rest and so diverged from all the rest of the regions. In the case of "we" words, on the other hand, it was children from North America who diverged from the rest of the population, as their rate of increase was significantly slower than that of the rest of the population.

Convergence and Divergence: WH Questions

The total usage of WH questions increased over time. Within the community, children from the Pacific Islands and Africa asked more WH questions than the others and children from the Middle East asked fewer. In addition, the Middle East and North America also increased faster in their questioning than the others. North America diverged with all groups except the Middle East, with whom it converges. The Middle East converges with all the others as well-Africa, Asia, Southeast Asia, Central and South Americas, and Caribbean and the Pacific Islands.

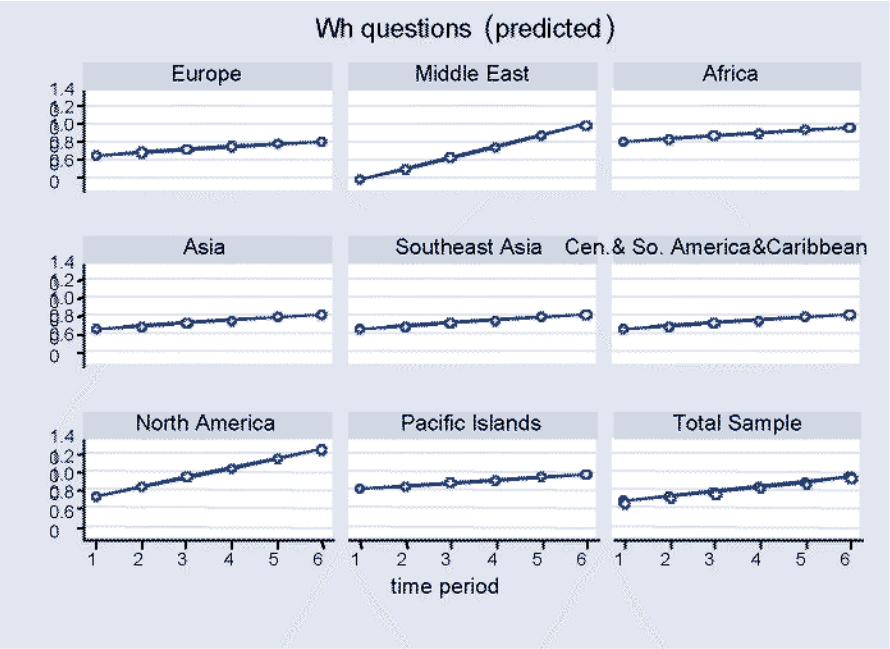


Figure 11. WH questions by region

Regional Dominance

In order to investigate how one of these groups might have influenced the others, we compared how changes within regions from each time period to the next affected the whole group's progress in the following two periods. To do this, we examined the change in a feature between T1 and T2 for children from each region and then checked to see what happened to that same feature in the total population

between T2 and T3. The region that was most closely associated with the entire group in the subsequent time period was labeled the "leading group."

Table 5 shows which region led the group for each feature in each time period. Looking at "affect," for example, what the overall group did between T2 and T3 was best predicted by what the children from the Middle East did between T1 and T2. The Middle East was then replaced as the leading group in the next period, T3-T4, by the children from the Central and South America and Caribbean group.

As evident from the chart below, different "leading groups" appeared in each time period for the eight features under examination. However, if we look at who led each feature, we see evidence to suggest that the group of children from Central and South America and the Caribbean led the way for the rest of the children in their use of both the singular and plural first person pronouns. In three out of four of the time period changes, this region preceded the larger population in both decreasing their use of "I" words and increasing their use of "we" words. No other region showed as much dominance over a group's use of a feature.

	Time interval			
Feature	**(1,2,3)**	**(2,3,4)**	**(3,4,)**	**(4,5,6)**
Affect	Middle East	C/S America & Caribbean	East Asia	Middle East & Pacific Islands
Apologies	Europe	Middle East	Africa	Pacific Islands
Future	Pacific islands	Middle East	Europe	East Asia
Hedges	Southeast Asia	Middle East, Pacific islands	Pacific Islands	Middle East
I	C/S America and Caribbean	C/S America & Caribbean	East Asia	Africa, C/S America & Caribbean
Junior Summit	East Asia	Southeast Asia, C/S America & Caribbean	Africa, Southeast Asia	C/S America & Caribbean
We	C/S America & Caribbean	C/S America & Caribbean	C/S America & Caribbean, N. America	East Asia
Wh questions	Africa	North America	Southeast Asia	C/S America & Caribbean

Table 5. Table of "leading" regions

Identity is not, of course, composed only of ethnic origin, and it is important to note that instances of entrainment to one linguistic style are also found for other groupings of the participants, for example by age and by gender. As reported in Tversky and Cassell (in preparation), girls use more emotion words than boys at the beginning of the forum, but boys' use of emotion words rises to meet girls' by the end of the first three months. Likewise, older children talk about the future more than do younger children. Yet by the end of three months, the younger children are referring to the future as often as do their older peers.

Language Choice Online

Another issue involving difference, and demonstrating regional dominance, concerns the use of English in the Junior Summit. Although automatic language translation was implemented for five languages—English, Spanish, French, Portuguese, and Chinese-the vast majority of messages were written in English. For participants who had the option of conversing in a local language or English, many started in a local language and then switched. Many interpretations of this pattern are possible. Some of the children told us they chose to write in English from the outset because it was an easy second language (as was the case for participants from India, for example), and they thought it would facilitate their communication with the other children. Certainly English is cited as the most widely-used language in inter-language situations (Durham, 2003).

The most common reason for using English reported by the participants was because they wanted to practice their English-in fact, "improving English skills" emerged as one of the more common benefits participants cited from the program. To take this explanation at face value, however, seems naïve. Why was the forum seen as a place to improve English? Other interviewees reported switching to English because most of the conversation was taking place in English. Why was most of the conversation taking place in English when native English speakers did not make up the majority of participants? From the chart below, it is apparent that as time wore on, participants wrote increasingly in English. During the first month online, 84% of the total messages were written in English while during the second and third months, it was 90% and 91% respectively.

		Number of messages posted				
	Individual participants	**Month 1**	**Month 2**	**Month 3**	**Total # of messages**	**Percent of total**
English	542	7232	13354	3630	23490	91.1%
Spanish	62	879	1124	214	2167	8.4%
French	62	145	60	21	224	<1%
Portuguese	20	286	264	102	635	2.5%
Chinese	10	38	62	4	99	<1%
Total	696	8580	14864	3971	25766	

Table 6. Messages posted in each language

Frustrations about language use were also reported by children interviewed five years after the summit. Several participants mentioned struggles they had or watched others have with the language translation mechanism. Two reported switching from Spanish to English in order to be better understood, though that did not always solve the problem for participants whose English was not strong.

Here it is clear that the Junior Summit community was constituted-and increasingly constituted over time-as an English-language venue. In fact, the children who were elected to be delegates to the Boston in-person event were almost exclusively fluent in English, so we can say that success in the Junior

Summit came with demonstrating one's proficiency in English. Notably, the issue of language choice remained in the realm of how the community was linguistically constituted by its participants, as opposed to who became or remained participants in the community. That is, rates of attrition varied regionally, but do not seem to be correlated to the use of English or another language. All regions experienced a loss of at least 50% of their original populations between the first month and the third. The percent of children online at the end the third month ranged from 52% among participants from South Asia, to just 20% among children from East Asia.

- Europe (49:20) = 41% remaining
- Middle East (25:12) = 48%
- Africa (24:8) = 33%
- East Asia (36:7) = 20%
- South Asia (42:22) = 52%
- Central and South America and the Caribbean (35:13) = 37%
- North America (47:20) = 43%
- Pacific Islands (25:11) = 44%

Participants' Views of Community Formation

Although all of our analyses rely on the children's own words for interpretation, it is important also to ask the children for their own thoughts about the nature of the Junior Summit. We therefore report here on some themes that emerged across the interviews carried out five years after the Junior Summit with participants from around the world.

In general terms, interviewees referred to the Junior Summit as a "big extended family" that was "united" and that made them feel like "we were in a massive global network of people all around the world."

> . . . it helped to, to learn that even though we are far away physically, there are a lot of ways to unite and our ideas and that we are very lucky to be in an age where we have the tools to communicate so easily and that, that when we communicate, even though we are from different cultures, we have something that make us the same. If we are young, we have the same idea, and we have the same spirit. I learned that from the junior summit. I think everyone who has participated, that even though we have different languages, different cultures, we live in different communities and different countries are different, [. . .] are different, religions, we have the same spirit. (Boy, 20, Argentina)

Finding One's Voice

One major issue addressed during the interviews was whether participants "felt heard" on the online forum, and by whom. The overwhelming majority of those interviewed (84%) said they felt their peers were listening to them.

> I guess, [can't think of an] email [I sent], and didn't get some sort of reply to. That sort of showed that someone was out there reading them and listening to what I was saying and so that no matter what you said, someone would come back with something about that subject, so you could tell that they really were listening. (Girl, 18, Australia)

For those children who said they did not feel heard, the main reasons cited related to adults in the world

not providing adequate feedback or up-take of their ideas. For example, one participant from Nepal was bothered when he found out that the ideas presented by the children at the in-person conference had not been implemented by governments around the world. Another participant from Botswana was disappointed by the response she got from her school when she presented a Junior Summit idea to her headmaster, who discouraged her from carrying it out.

Interested in the relationship between commitment to building the forum community, and commitment to being listened to, we divided the 78 interviewees into two groups based on how often they posted to the forum. An equal percentage of both populations reported "feeling heard" by their peers during the online forum. However, differences appeared when we looked at who else it was important to have listening to them. High posters also said it was important to be heard by world leaders, at a rate three times that of low posters (33%). Indeed, several of our findings point to the interpretation that greater investment in the program (as measured by number of posts) was correlated with higher expectations of outcomes of the program. While periodic contributors may have been content with feeling heard by their peers, those who put more into the program wanted powerful adults listening as well. Unfortunately, those who did put more into the forum may also have ended up more disappointed, since the truth was that peers were more likely to listen than were powerful adults outside the forum. Ultimately, the Junior Summit participants, brought together to become a community to effect change in the world, were more effective at constituting themselves as a community than as agents of global change. Given that so many of the children reported community affiliation, support, and communication as benefits of the Junior Summit, it is not surprising that those participants whose goals for the summit were action rather than community did not have their goals entirely fulfilled.

Many mentioned the link between voicing opinions and confidence gained from the Junior Summit. Gaining confidence in their opinions and assertiveness in voicing those opinions were the most common attitude changes that participants reported as an outcome of the forum.

> And most importantly [the Junior Summit] gave me the confidence that people would listen to me if I have something positive to contribute. I think that was a very important thing. Confidence is something that cannot be really quantified but it can make a wonderful difference to [a] personality. (Boy, 17, India)

> I think before I went to the Summit, I was not so vocal. I didn't really believe what I said.... And from there, it's like they gave everybody a chance to participate and then you could see that whatever you say, whatever your views are, the way I really [felt?], and it made you more confident and in that way you believed in yourself every time you talked. And this is something I've carried with me since then. (Girl, 20, Botswana)

For the participants interviewed, the community was described as empowering-as support for individual performance.

> I mean I felt really powerful. I felt we could do, I mean I couldn't do anything by myself at all but with people, with everyone there like with the help of adults but with all the good will that was there, with this great environment and this all this tech instrument we could do anything. (Boy, 19, Mexico)

> ...You know, the most important part in where I found my motivation in all this, or encouragement was to look at other young people in different countries doing so much more for their society and then looking back over here and seeing that, oh, the youngsters here are not really interested in doing things, you know? So, that was where I got my inspiration. OK, if someone in U.S. can do something, as amazing as this, or someone in Nigeria can do something like this, then why can't I not do that in my society over here where I need to do it. That was the most important, or inspirational part in motivating part which I got out of it, and why I kept on doing what I wanted to do for the society. (Girl, 19, India)

While many participants mentioned feeling empowered and inspired to be social activists, this identification as an activist was facilitated not by their contact with the 20 adult moderators of the forum--many of whom had spent lifetimes as social activists-but through interaction with their peers and by the community they had created. As one delegate from Morocco put it, "Well, the summit made it easier to change the world, obviously, because we had 3,000 kids supporting us, you know. We're going to do it, and we felt like our voices were louder."

Friendship

Twenty-three of the thirty-seven young people interviewed reported making real friends (friends like the friends they had in their local communities) during the Junior Summit, while only eight said they had not. Many of the friendships formed at the Junior Summit have lasted until today. In some cases, those bonds replaced community support missing from local environments.

> I suppose Junior Summit was really an amazing experience in meeting a whole bunch of people who were very much like myself in some respects, and …So it was a great, and that also continued on, like those friendships continued on for like three years and were a significant part of my life for those next three years or four years, and still are today, in some respects. I suppose, to reflect on this a bit more, …[in] grade 10, there weren't like as many of my friends in my school anymore, so to some extent, for about a year, like the Junior Summit friends almost became my friends, for like, well, didn't really become my friends, but they filled a gap that I had for like a year, and I think that that kind of dropped away as [. . .] within a year of junior summit, but it was a good oppor-, like a good timing, to some extent, for me. (Boy, 20, Australia)

In many cases, the friendships moved offline. A participant from Malta explained her relationship with a girl from Argentina by saying:

> Even though we never met, we only saw [pictures of] each other, we never even phoned each other or heard our voices or anything, but we sent a lot of pictures to each other. …once, [. . .]I didn't have Internet for a couple of months. I had no Internet subscription. I didn't pay for it. And because she didn't hear from me she was really panicking, and I think she called me then…She panicked that something happened to me. And I was really amazed. I said, oh, she's a real friend.

Friendships from the Junior Summit extended not just offline but also beyond the children. In one instance, the mothers of two participants, one from Pakistan and one from India, developed a friendship

online and continued to communicate with one another for several years, in an era when tensions between those two countries ran high.

Conclusion

Word frequency features of the kind we have described in this study have been found to be reliable predictors of demographic variables such as age and gender. Here we have used word frequencies and content analyses to discover the ways in which young people from very different cultural, linguistic, and socio-economic backgrounds increasingly referred to themselves as a community, speaking in the collective voice, and concurring on the topics of conversation, the goals of the community, and their strategies for achieving them, through interacting in an online forum.

Research online and offline suggests that we match our language to that of our conversation partners. A number of researchers have examined how participants in conversation exhibit language synchrony or entrainment on the conversational level, turn-by-turn level, and level of lexical items (Brennan, 1996; Niederhoffer & Pennebaker, 2002). The field of sociolinguistics has looked at how sharing ways of talking can in and of itself constitute social action-the construction of a group identity. For young people, such as those of the Junior Summit, identity construction is a careful negotiation between self and community in which linguistic style plays a key role (Eckert, 1996). Here we have demonstrated significant convergence in and mutual influence over language use during as short a period of time as the first three months of the online forum of which these young people were a part. Interestingly, as the community came together as one, the participants reported that their appreciation for diversity, their ability to see different perspectives, and their positive reactions to one another increased. They began to see each other as friends, and to care about what was happening in the parts of the world that their new friends came from.

Of course, convergence of the sort illustrated here could be perceived as an essential part of the formation of a new international community, or as an index of the loss of local culture. In this context, the participants' use of English points up an important aspect of how language does not only reflect, but also *constitutes* community. Each of the young people who participated in the Junior Summit is already a member (albeit sometimes a liminal member) of a nation state, with attendant ideologies and constructions of self and community. The organizers of the Junior Summit invited the children to chart a new ideological and practical space. The children took on that challenge. In some ways their communication reflected the world in which their parents live-speaking English correlates with success on the international stage, for example-and in some ways they struck off in new directions-flaming was rare, and giving and requesting feedback was valued.

Finally, we have eschewed measuring the Junior Summit against external definitions of "authentic community" in the "real world," for two reasons: because we wished to hear the style in which the children themselves imagined their community (Anderson, 1991), but also because we do not believe that the online world is so discontinuous with the real world. The processes described here of dynamic and cross-cutting constructions of self and community within this extremely diverse population are not an outgrowth of new technology. Technologies such as online virtual community software are, rather, a

miraculous lens through which to watch community constitituting of the kind that happens every day.

Acknowledgments

Thanks to Jacqueline Karaaslanian for all her help with every stage of the Junior Summit from its inception to its evaluation, to Modupe Adeleye, Hangyul Chung, and Megan Tucker for their extraordinarily insightful and painstaking coding and analysis, to Jenya Kaganova for her talented statistical consultation, Ginny Dorne for unbelievably complex travel arrangements, to the Kellogg Foundation for gracious and generous funding, and to the 3062 Junior Summit participants who have illuminated our vision of what it means to be a child, and a citizen of the world, and have made our lives immeasurably richer.

References

Anderson, B. (1991). *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (revised ed.). London: Verso.

Bales, R. F. (1951). *Interaction Process Analysis*. Cambridge: Addison Wesley.

Brennan, S. E. (1996). Lexical entrainment in spontaneous dialog. Proceedings, 1996 *International Symposium on Spoken Dialogue*, ISSD-96, 41-44.

Cassell, J. (2002). "We have these rules inside": The effects of exercising voice in a children's online forum. In S. Calvert, R. Cocking & A. Jordan (Eds.), *Children in the Digital Age* (pp. 123-144). New York: Praeger Press.

Climent, S., Moré, J., Oliver, A., Salvatierra, M., Sànchez, I., Taulé, M., & Vallmanya, L. (2003). Bilingual newsgroups in Catalonia: A challenge for machie translation. *Journal of Computer-Mediated Communication, 9* (1). Retrieved December 28, 2004 from http://jcmc.indiana.edu/vol9/issue1/climent.html

Durham, M. (2003). Language choice on a Swiss mailing list. *Journal of Computer-Mediated Communication, 9* (1). Retrieved December 28, 2004 from http://jcmc.indiana.edu/vol9/issue1/durham.html

Eckert, P. (1996). Vowels and nail polish: The emergence of linguistic style in the preadolescent heterosexual marketplace. In N. Warner, J. Ahlers, L. Bilmes, M. Oliver, S. Wertheim & M. Chen (Eds.), *Gender and Belief Systems: Proceedings of the Fourth Berkeley Women and Language Conference* (pp. 183-190). Berkeley: Berkeley Women and Language Group.

Eckert, P., & McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology, 21,* 461-490.

Flanders, N. (1970). *Analyzing Teacher Behavior*. Reading, MA: Addison Wesley.

Herring, S. C. (1996). Two variants of an electronic message schema. In S. C. Herring (Ed.), *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives* (pp. 81-108). Amsterdam: John Benjamins.

Herring, S. C. (2004). Computer-mediated discourse analysis: An approach to researching online behavior. In S. Barab, R. Kling, & J. H. Gray (Eds.), *Designing for Virtual Communities in the Service of Learning*. New York: Cambridge University Press.

Hofstede, G. (1980). *Culture's Consequences: International Differences in Work-Related Values*. Newbury Park, CA: Sage.

Jones, Q. (1997). Virtual-communities, virtual settlements & cyber-archaeology: A theoretical outline. *Journal of Computer-Mediated Communication, 3* (3). Retrieved December 28, 2004 from http://jcmc.indiana.edu/vol3/issue3/jones.html

Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., & Crawford, A. (2002). Internet paradox revisited. *Journal of Social Issues, 58* (1), 49-74.

Kraut, R., Lundmark, V., Patterson, M., Kiesler, S., Mukopadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist, 53* (9), 1017-1031.

Lakoff, R. T. (1975). *Language and Woman's Place*. New York: Harper Colophon Books.

Ma, R. (1996). Computer-mediated conversations as a new dimension of intercultural communication between East Asian and North American college students. In S. C. Herring (Ed.), *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives* (pp. 173-186). Philadelphia: John Benjamins.

Meagher, M. E., & Castanos, F. (1996). Perceptions of American culture: The impact of an electronically mediated cultural exchange program of Mexican high school students. In S. C. Herring (Ed.), *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives* (pp. 187-202). Philadelphia: John Benjamins.

Murdock, G. P. (1945). The common denominator of cultures. In R. Linton (Ed.), *The Science of Man in the World Crisis* (pp. 123-142). New York: Columbia University Press.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology, 21* (4), 337-360.

Palfreyman, D., & Khalil, M. A. (2003). "A funky language for teenzz to use": Representing Gulf Arabic in instant messaging. *Journal of Computer-Mediated Communication, 9* (1). Retrieved December 28, 2004 from http://jcmc.indiana.edu/vol9/issue1/palfreyman.html

Panyametheekul, S., & Herring, S. C. (2003). Gender and turn allocation in a Thai chat room. *Journal of Comuter Mediated Communication, 9* (1). Retrieved December 28, 2004 from http://jcmc.indiana.edu/vol9/issue1/panya_herring.html

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. Mahwah, NJ: Erlbaum Publishers.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54,* 547-577.

Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the lifespan. *Journal of Personality & Social Psychology, 85* (2), 291-301.

Rafaeli, S., & Sudweeks, F. (1997). Networked interactivity. *Journal of Computer-Mediated Communication, 2* (4). Retrieved December 28, 2004 from http://jcmc.indiana.edu/vol2/issue4/rafaeli.sudweeks.html

Rheingold, H. (1994). *The virtual community in a computerized world*. London: Secker & Warburg.

Rourke, L., & Anderson, T. (2002). Exploring social interaction in computer conferencing. *Journal of Interactive Learning Research, 13* (3), 257-273.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (1999). Assessing social presence in asynchronous, text-based computer conferencing. *Journal of Distance Education, 14* (3), 51-70.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological issues in the content analysis of computer conference transcripts. *International Journal of Artificial Intelligence in Education, 12* (1), 8-22.

Sinclair, J. M., & Coulthard, R. M. (1975). *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*. London: Oxford University Press.

Strauss, A. L., & Corbin, J. (1994). Grounded theory methodology—an overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of Qualitative Research* (pp. 273-285). Thousand Oaks: Sage Publications.

Subrahmanyam, K., Kraut, R., Greenfield, P. M., & Gross, E. (2000). The impact of home computer use on children's activities and development. *The Future of Children: Children and Computer Technology, 10* (2), 123-144.

Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychological Review, 96* (3), 506-520.

Tversky, D., & Cassell, J. (in preparation). *How to win a world election: Gender and leadership in an online international youth forum.* Northwestern University.

Wellman, B. (2001). Computer networks as social networks. *Science, 293* (5537), 2031-2034.

Wellman, B., Boase, J., & Chen, W. (2002). The networked nature of community: Online and offline. *IT & Society, 1* (1), 151-165.

About the Authors

Justine Cassell is Professor of Communication Studies and Computer Science at Northwestern University, and Director of the Technology and Social Behavior track of the graduate program in Media, Technology and Society. Before coming to Northwestern, Cassell directed the Gesture and Narrative Language Research Group at the MIT Media Laboratory. She holds undergraduate degrees in Comparative Literature from Dartmouth and in Lettres Modernes from the Université de Besançon (France), an M.Phil in

Linguistics from the University of Edinburgh (Scotland), and a double Ph.D. from the University of Chicago in Linguistics and Psychology.
Address: Media, Technology & Society, Northwestern University, 2240 Campus Drive, 2-148 Evanston, IL 60208 USA

Dona Tversky is a Research Scientist at the MIT Media Laboratory where she directs the follow-up study of the Junior Summit '98 program. Prior to coming to MIT, Tversky completed a Masters of Public Health at the University of the Western Cape in South Africa, and conducted research on ways to use media for public health promotion. She holds an undergraduate degree from Yale University.
Address: 924 Spruce Street Apt.102, Philadelphia, PA 19107 USA

© 2005 Journal of Computer-Mediated Communication