

Paper

A Virtual Barbeque: A Corpus Linguistics Approach to Studying an Emergent Community.

Roz Horton

University of Manchester

hortonr@cs.man.ac.uk

Richard Giordano

University of Manchester and Brown University

Richard_Giordano@brown.edu

Keywords: corpus linguistics, Internet, electronic community

The background

The concept of community in an online context is a relatively recent phenomenon, and deals with social groupings formed through the shared medium of computer mediated communication (CMC). In addition, the civic metaphor (for instance, the 'virtual community') is coming to replace the information superhighway as the dominant metaphor for understanding and constructing cyberspace. As a concept centrally related to the civic metaphor, it is therefore important to understand what we mean by 'virtual community', a expression much used in varying contexts to encompass disparate social groupings. There is, consequently, a need to identify the constituent features that differentiate between genuine online community (which, we argue, involves networks of personal relationships) and other social aggregations.

Recent research into cultural formations in cyberspace proposes different ways of identifying virtual community and considers the similarities and differences between virtual (online) community and embodied 'real space' community. For instance, Bruckman (1992) considers that technological features of the virtual environment combine with self-selected membership to create a community with a strong shared sense of values. She also argues that shared activity reinforces community, a theme which is developed in her later work on her created professional community environment, MediaMOO (Bruckman and Resnick, 1995). Reid (1994) has studied interaction on MUDs and MOOs, which are text-based virtual reality systems. She identifies users of these systems as a distinct cultural group which is characterized by their use of novel methods of textualizing non-verbal communication. Again, she considers that the group has its own distinct systems of meaning. A central issue for Reid is an examination of forms of social and technological control which are employed to regulate interaction and penalize or exclude disruptive influences. Smith (1992) evaluates virtual community in terms of its capacity to create collective goods in the form of the provision of social networks, the production of 'knowledge capital' and communion; this last term is understood as a sense of membership, which is fostered by personal and emotional communication. Smith is also concerned with mechanisms of control which can be used to overcome obstacles to community formation and deal with violations of community standards. Rheingold (1995) has proposed a definition of virtual communities, "social aggregations that emerge from the Net when enough people carry on those public discussions long enough, *with sufficient human feeling*, [emphasis added] to form webs of personal relationships in cyberspace." Both this prerequisite of communicated, human feeling, and the concept of a shared system of meanings and/or values, are key factors in identifying the existence of virtual community as opposed to any other sort of group or aggregation.

The researchers mentioned above have, in general, taken an ethnographic and qualitative approach to the communities and cultural formations under study. As language is *the* medium of computer-mediated communication, we argue that empirical linguistic analysis should be both an alternative and fruitful way to understand the emergence and structure of a virtual community. A 'speech community' can potentially be identified by linguistic convergence at a lexical and/or structural level. Because Computer Mediated communication, as we suggest, is strongly oral in nature (December, 1993) (Ferrara et al., 1991), we believe that speech accommodation theory (Giles and Powesland, 1975) is an appropriate model by which to explain the acquisition of shared meanings, being the first of the two hypothesized major elements of virtual community.

The practical study

The corpus studied consists of eight months' postings to one electronic discussion list consisting of 2,600 messages. Messages are gathered, organized and mailed daily to members of the list. The discussion list is composed of people who have little else in common except for the fact that they listen to a certain radio programme in New York City, or obtain tapes of the show if they reside elsewhere. The show they listen to is Vin Scelsa's "Idiots Delight," and the name of the mailing list is the Idiots Delight Digest (IDD). Participants to the mailing list reside in most areas of the United States, including at least a half-dozen countries outside North America, although most are concentrated in the New York-Northern New Jersey region. We gathered postings from the very first issue of the digest, through to when they organized their first face-to-face barbeque. Until the time of the barbeque, most members would not recognize each other on the street. The members explicitly decided that they would not limit their contributions to Vin Scelsa, Idiots Delight, music, or any other topic. As one member posted, "This is a virtual barbeque, and we should talk about anything that's on our minds. That's what a barbeque is there for." Consequently, some postings contain very personal messages, like news of the death of a relative, marital breakdowns, the loss of employment, alcoholism and wife abuse.

The corpus has been pre-processed and analysed using everyday off-the-shelf tools such as Microsoft Word and Excel, to determine the profile and core membership of the hypothesized community. An important aim of this work has been to develop a methodology that could be applied to other corpora, using computing tools which are commonly available and quickly learned by a researcher without a computing background.

The methodology is strongly influenced by Stubbs's (1996) approach to corpus analysis in the British, neo-Firthian tradition. Stubbs proposes a model of meaning which is situated in the relationship between text, writer and reader; meaning is analysed distributionally on the basis of observed, objective textual evidence. Specifically, to paraphrase Firth, the meaning of a word can be deduced from the company it keeps--not only is meaning conveyed directly, but also indirectly through patterns of co-occurrence of words. These lexical collocates are not easily observable except through computer-assisted corpus analysis. Stubbs shows how corpus evidence such as the patterns of usage of keywords/focal words (Williams, 1976) (Firth, 1935) help to explain a linguistic structure that transmits and reinforces culture.

We are using TACT, the concordance program, to carry out a surface linguistic analysis of the corpus seeking evidence that tests the definition of a virtual community as a social aggregation in cyberspace which possesses a flexible but characteristic set of shared meanings, located in the speech community in the form of a consensus interpretation, and that meanings within the community differ significantly from the equivalent meaning possessed by the wider culture in which the virtual community is embedded. Further, members of that community communicate with a sufficient degree of human feeling to create and maintain a sense of communion and shared presence. In addition, mechanisms of control, including sanctions, are used to regulate social interaction and shared social activities. These sanctions indicate shared normative behavior.

The form of evidence we seek are patterns of usage and collocations of selected keywords in context which can be used to identify the hypothesized community traits. These keywords are to be initially derived as follows:

1. **Shared meaning:** Stubbs (1996:157ff) discusses keywords which have been used in various studies to identify, by their usage, elements of transmitted culture. These keywords and their collocates are examined in context to find out whether patterns of usage differ significantly from the wider cultural

usage, thus perhaps indicating the development of a characteristic community consensus meaning. Other potential keywords are identified by manual inspection of the corpus.

2. **Human feeling:** an initial keyword-list of personal/emotional words has been drawn up, and further relevant keywords will be added from both examination of tables of collocates of the initial keywords, and manual inspection of the corpus.

Other features we examine include evidence of lexical convergence; the handling of conflict and the building of a group topic consensus; and the influence of individual members' language on the group's language related to the frequency of their contributions.

Further analysis.

We are currently conducting a linguistic analysis on the corpus, and we will report to the meeting our findings about community formation, and in particular the findings, will be compared with the work of Smith, Reid and Bruckman. A contra-indication of community formation will also be discussed: the formation of what Beniger (1987) calls a *pseudo-community* which occurs when mass media is personalized such that the recipient believes that a communication is meant for themselves alone when in fact it is aimed at a much larger set of recipients.

Our preliminary investigation indicates that Stubbs's methodology is useful in discovering and measuring the emergence and growth of a virtual community. The number and extent of shared meanings appears to grow over time, and the emotional cohesion of the group has increased, as well. We also believe that our configuration of off-the-shelf tools, available to most people involved with humanities computing, will be both relevant and useful.

References

- Anderson B., (1993) "Imagined Communities" in Anderson B.(ed.) *Cyberspace: First Steps* (Cambridge, Mass.: MIT Press)
- Baron N. S., (1984) "Computer-mediated communication as a force in language change" in *Visible Language* Vol XVIII (2), pp 118-141.
- Baym N., (1995) "The emergence of Community in Computer-Mediated Communication" in Jones S. (ed.) *Cybersociety: Computer-Mediated Communication and Society* (Thousand Oaks: SAGE Publications)
- Beniger J.R., (1987) "Personalization of Mass Media and the Growth of Pseudo-Community" in *Communication Research* Vol. 14 No. 3 pp. 357-371.
- Bolinger D., (1975) *Aspects of language* (New York: Harcourt Brace Jovanovich)
- Bruckman A., (1992) "Identity Workshops: Emergent Social and Psychological Phenomena in Text-Based Virtual Reality" Master's thesis, MIT Media Laboratory, 1992
- Bruckman A., (1995?) "Finding One's Own in Cyberspace" (??)
- Bruckman A. and Resnick M., (1995) "The MediaMOO Project: Constructionism and Professional Community" in *Convergence* 1:1.
- Cohen A. P., (1985) *The Symbolic Construction of Community* (London: Tavistock Publications Ltd)
- Coulthard M.(ed.), (1994) *Advances in Written Text Analysis* (London: Routledge)
- Coulthard M.(ed.), (1992) *Advances in Spoken Discourse Analysis* (London: Routledge)
- December J., (1993) "Characteristics of oral culture in discourse on the net" Paper presented at the twelfth annual Penn State Conference on Rhetoric and Composition.

- Ferrara K., Brunner H. and Whittemore G., (1991) "Interactive written discourse as an emergent register" in *Written Communication* Vol. 8 (1) pp. 8-34.
- Giles H. and Powesland P. F., (1975) *Speech Style and Social Evaluation* (London: Academic Press Inc.)
- Halliday M.A.K., (1985) *Spoken and Written Language* (Oxford: OUP)
- Harrison T. M., and Stephen T. D., "The Electronic Journal As the Heart of An Online Scholarly Community" in *Library Trends* 1995 Vol. 43 No. 4 pp. 592-608.
- Lakoff R. T., (1982) "Some of my favorite writers are literate: the mingling of oral and literate strategies in written communication" in D.Tannen (ed.) *Spoken and Written Language: Exploring Orality and Literacy* pp. 239-260 (Norwood, NJ: Ablex)
- Lea M. and Spears R., (1995) "Love at First Byte? Building Personal Relationships Over Computer Networks" in Wood, J.T. and Duck, S (eds.) *Under-Studied Relationships: Off the Beaten Track* (Thousand Oaks: Sage Publications)
- Lea M. and Giordano R., (1996) "Representations of the group and group processes in CSCW research: A case of premature closure?" (in press)
- Mott P., (1995) "Towards a Winograd/Flores Semantics" in *Minds and Machines* 1995 Vol. 5 No. 1 pp. 69-87.
- Oldenburg R., (1989) *The Great Good Place*, (New York: Paragon House)
- Ong W. J., (1982) *Orality and Literacy: The Technologizing of the Word* (London: Methuen)
- Reid E., (1994) "Cultural Formations in Text-based Virtual Realities "Masters thesis, Department of English, University of Melbourne
- Rheingold H., (1994) *Virtual Community: Finding Connection in a Computerized World* (London: Minerva)
- Shank G., (1993, March 22) "Abductive multiloguing: the semiotic dynamics of navigating the net" in *The Arachnet Electronic Journal on Virtual Culture* 1 (1).
- Smith M. A., (1992) "Voices from the WELL: The Logic of the Virtual Commons" Masters thesis, Department of Sociology, UCLA
- Spitzer M., (1986) "Writing style in computer conferences" in *IEEE Transactions on Professional Communication* Vol. 29 (1) pp. 19-22.
- Stubbs M., (1996) *Text and Corpus Analysis: Computer-assisted Studies of Language and Culture* (Oxford: Blackwell)
- Tannen D., (1982) "The Oral/Literate Continuum in Discourse "in D.Tannen (ed.) *Spoken and Written Language: Exploring Orality and Literacy* pp. 1-16 (Norwood, NJ: Ablex)
- Tönnies F., (1963) *Community and Society* (New York: Harper & Row)