# English in computer-mediated environments: a neglected dimension in large English corpus compilation

*Wengao Gong*
Department of English Language and Literature
National University of Singapore
g0402711@nus.edu.sg

## 1. Introduction

Language and language use can often be influenced and constrained (sometimes even shaped) by new technologies. Human history has witnessed this happen with printing technology, broadcasting technology, telephone technology, television technology, and more recently computer and Internet technology. Radio, for instance, has brought a new kind of language which quickly yielded several subvarieties such as commentary, news, and weather. The advent of television added a further dimension, which similarly evolved new subvarieties (Crystal, 2001: 225). These new audio and visual technologies also brought about the blending of spoken and written language, creating new categories such as scripted speeches which are written to be heard. The advent of computers and the Internet has created a new linguistic dimension: computer-mediated language or Netspeak in Crystal's terms (2001). According to Crystal, Netspeak is "not simply a new variety of English, but a whole new medium, comparable to speech and writing in its distinctiveness and generality, and subsuming a great deal of linguistic variation" (2003: 426).

As a new frontier of linguistic investigation, computer-mediated English or Netspeak has attracted the attention of quite a number of researchers. Many articles concerning the genres of Internet-based language use and the linguistic features of some of these genres have been published over the past 10 years. Despite that, the importance of computer-mediated English has not been fully recognized. An important piece of evidence is that there is not even one large corpus of English for general purpose which consists of a balanced computer-mediated English component. While this may well be attributed to the slow updating process of large corpus compilation, it is also highly likely that computer-mediated English has not yet been recognized as an established form of language use in spite of its popularity. As a newly emerged dimension, computer-mediated English needs further and more systematic investigation. This paper aims to illustrate why English in computer-mediated environments as a whole should be included in large English corpora and which specific genres should be included. It will also discuss the necessity of constructing an international corpus of computer-mediated English.

## 2. English in computer-mediated environments: why is it important?

The rapid development and easy availability of information and communication technology have contributed considerably to the flourish of computer-mediated communication (CMC). According to Herring (1996:1) CMC is communication that takes place between human beings via the instrumentality of computers. Ooi (2002: 91) redefines CMC with more emphasis on the multi-modal nature of the medium by calling it "a mode of human communication that centrally involves the computer as the medium, and made via a hybrid of speech, writing, graphics and orthography." In fact, the term "computer-mediated communication (CMC)" is sometimes used to refer to different things. In its broader sense, it refers to all communication activities mediated by computer networks. Examples of CMC in this sense include e-mail, listserv mailing lists, Usenet groups, Internet Relay Chat (chatroom), social MUDs, web pages (weblogs), ICQ, MSN, Skype, audio and video chat, graphical Virtual Reality environments, SMS via mobile phones, etc. CMC in its narrower sense often refers to text-based synchronous and asynchronous communication where participants interact by typing a message on the keyboard of one computer which is read by others on their computer screens. In order to avoid ambiguity, the author of this paper uses the term "computer-mediated environments" to refer to CMC in its generic sense while reserves the original term "CMC" for its narrower sense.

Computer-mediated environments can be divided into different modes according to the medium used. Affected by the technological affordance of the medium or mode and the difference in situation and function, English used in different CMC mode takes on different linguistic features. English in computer-mediated environments can be roughly classified into five situations: English on web pages, English in asynchronous settings, English in chatroom, and English in blogs (weblogs).

### 2.1 English on web pages

Anything that can exist as a computer file can be made available as a Web document: text, graphics, sound, video, etc. Thus, when we are talking about language on web pages in general and English on web pages in particular, we are not talking about the written language only. Nevertheless, existing research regarding language (English) on web pages focuses more on the written aspect of the language used. Anything that has been written can, in principle appear, on the Web, thus even a tiny exposure to the web demonstrates its linguistic range. Yet, nobody has ever carried out a detailed and systematic investigation about the linguistic features of English on web pages and how it might be different from the English used in conventional media. Almost all of the existing research findings about web language are concerned with its superficial features such as its graphic linguistic existence, its information structure, its sentence or paragraph lengths, its interactivity, etc. Crystal (2001) gives a rather detailed discussion about all these aspects. According to him, texts on the web are displayed in both interrupted linear format and non-linear format. The former can be read linearly just like conventional print materials while the latter can be read in a multi-dimensional way. In addition, the Web is graphically more eclectic than any domain of written language in the real world. Whatever the variety of written language we have encountered in the paper-based world; its linguistic features have their electronic equivalent on the Web (Crystal, 2001: 197). Due to the fact that the screen is often divided into many functional areas, the on-screen textual description of each area tends to be short. This feature of the Web has contributed a great deal to the short sentences and short paragraphs of web texts. This finding is echoed by Ide, et al. (2002: 844), as can be seen from the following quotation:

> Texts drawn from the web exhibit characteristics that are similar, but not identical, to other text types, suggesting that they can be regarded as falling into a genre of their own. In particular, written web materials contain dense, information-packed language that is also found in official documents and academic prose. However, they also appear to be more cryptic and terse, containing shorter paragraphs than those found in paper-based materials.

One thing of particular interest to researchers studying the Web pages is the hypertext link, which according to Crystal (2001) is the most fundamental structural property of the web, without which the medium would not exist. Its non-linear, non-sequential, non-hierarchical and multimodal (employing images, sound and symbols as well as text) nature has placed it in stark contrast to traditional printed texts (Macfadyen and Doff, 2003: 4). Burbules (1997) also considers the *hyperlink* as the key feature of texts on the Web, and explores some of the different roles links may play beyond their simple technical role as shortcut: interpretive symbol for readers, bearer of the author's implicit ideational connections, indicator of new juxtapositions of ideas (cited in Macfadyen and Doff, 2003: 6). Moreover, texts on web pages are not mono-semiotic (i.e., purely represented by characters) but rather multi-semiotic (represented by a combination of characters, signs, symbols, color schemes, etc. which are either static or animated). Whether this feature has any influence on the language of web pages is yet to be found out.

Ide, et al. (2002) have carried out an experimental investigation about the linguistic features of American English on the Web using a small corpus (following the criteria used in data collection for American National Corpus) which they created out of texts produced by Americans on US-based websites. They found that “in general, texts taken from the web represent a particular type of prose—in particular, a formalized, dense type of prose characteristic of formal documents” (2002: 842). In spite of the similarities found among the web written texts and paper-based texts, they hold that materials produced for the web would not exhibit characteristics of informal or even argumentative prose. They argue that the Web is not a source of the range of written texts that readers frequently encounter. As such, web texts lack the variety and distribution of linguistic features that can be found in many texts. Therefore, texts from the Web alone are not enough for constructing a representative and balanced corpus. Their findings may be skewed by the data they have used. Due to many practical considerations, they have only included texts from gov and edu sites into their corpus. Both types of websites are places where formal texts are more likely to be located. While admitting their drawbacks in using limited range of websites, they still conclude that web-based texts are only representative of a small slice of the range of genres encountered by human readers everyday, and therefore cannot be used to provide a comprehensive view of American English.

## 2.2 English in asynchronous CMC

English in asynchronous computer-mediated communication is characterized by the language used in email, and other communication modes which are based on emails, for instance, bulletin board system (BBS), and listserv mailing lists. Two main features of this kind of communication may have shaped their linguistic features: asynchronicity and interactivity. The former allows participants more time to plan and revise their messages if they like. On the whole, language in asynchronous CMC is linguistically more complex than that in synchronous mode. The easy-to-use nature of email system has made mail replies very convenient, thus increasing the interactivity of this mode of communication, which in turn contributes to the formation of the dialogic character of e-messaging (Crystal, 2001). According to Crystal’s research, the length of the text comprising the body of an email is relatively short: the vast majority fitted easily into a single screen view. Emails from institutions were much longer than private ones. The paragraph structure of the body text is also short. The kind of language used in email is often closely related to the social distance between two communicators and the purpose of communication. In spite of that, email tends to be less formal as other edited forms of writing. One reason is that emails are supposed to fulfil less formal purposes and the other is “the relative openness of email as a new communication mode that has not yet been colonized by rigid prescriptive norms" (Herring, 2001: 618).

As far as the discourse features are concerned, the language of asynchronous messaging is a curious mixture of informal letter and essay, of spoken monologue and dialogue. At the same time, it lacks some of the most fundamental properties of conversation, such as turn-taking, floor-taking, and adjacency pair (Crystal, 2001: 148)

Quite a number of existing studies are concerned with email and email-based BBS discourse and the focus of these studies is often on aspects like linguistic complexity, interactional patterns, and the length of messages. Again, no systematic comparison has been made between conventional letters and emails. Hard evidences concerning the similarities and differences between these two modes of communication are yet to be found.

## 2.4 English in Chatroom

With the popularity of computer-mediated communication in general and Internet Relay Chat (IRC) (or chatroom in non-technical terms) in particular, many people have started conducting conversation in written form. This novel practice is mainly attributed to the widespread use of computers and the Internet. Of course, this is not the first time for technology to bring about drastic changes to communication forms. The advent of telephone, for instance, has brought about considerable changes to the way people interact with each other and at the same time changed our conception about what conversation should be like. Due to the telephone's communicative affordances (that is, what the telephone can do and what it cannot), telephonic conversation has taken on several different features from face-to-face conversation. One main reason is that "the telephone places speakers in a situation of 'cuelessness': that is there is no recourse to the non-verbal cues that can be relied on in situations of physical co-presence" (Hutchby, 2001: 86). The telephone technology has broken through the constraints of space and time on human interaction, but as a compromise it blocks some very important channels (mainly paralinguistic ones such as gestures, postures, and eye contact) which people rely heavily on in face-to-face communication settings. As a result, when people are conducting telephonic conversations, they will have to use certain strategies to compensate for the cuelessness, which in turn contribute to the formation of some unique features. One of these features is that telephone conversation displays an opening pattern involving a summons-answer sequence which is not likely to be found in face-to-face conversations (for details, see Schegloff, 1968).

Chatroom conversation, due to its nature of being computer-mediated and text-based, places "speakers" in a situation of greater 'cuelessness' than that of telephone conversation. In this case, almost all paralinguistic and non-verbal channels have been blocked; the only means left is the written language itself. Besides, a majority of chatrooms are operated under the so-called "one-way transmission protocols" (Herring, 1999). This one-way system only allows the "speaker" (i.e., the message sender) to send his or her message in its entirety; therefore, the "hearer" (i.e. the potential message receiver) knows nothing about what message is being constructed before it is displayed on the monitor screen. This is drastically different from face-to-face conversation and telephone conversation where the hearer is able to monitor what the speaker is saying and decide when to contribute. Besides, the synchronous nature of chatroom conversation imposes certain temporal constraints on "speakers". All these constraints have contributed to shaping the unique features of chatroom conversation. Some of these features are: dominant use of monosyllabic words, frequent overlaps, lacking conventional adjacency pairs, grammar chiefly characterized by highly colloquial constructions and non-standard usage, nonce-formations, heavy use of non-standard formations, jargon, and slang for affirming group identity, playing with language, and so on (Crystal, 2001).

The following quotation from David Crystal sufficiently illustrates why English in chatroom situation is worthy of investigation.

> From a linguistic point of view, I find chatgroup language fascinating for two reasons. First, it provides a domain in which we can see written language in its most primitive state. Almost all the written language we read has been interfered with in some way before it reaches us—by editors, subeditors, revisers, censors, expurgators, copy-enhancers, and others. Chatgroups are the nearest we are likely to get to seeing writing in its spontaneous, unedited, naked state. Secondly, I see chatgroups as providing evidence of the remarkable linguistic versatility that exists within ordinary people—especially ordinary young people (Crystal, 2001: 169-170).

### 2.5 English in weblogs

Weblogs (blogs) defined by Herring et al. (2005:1) as “frequently modified web pages in which dated entries are listed in reverse chronological sequence” are becoming an increasingly popular form of communication on the World Wide Web. High expectations for social functions have been placed on this new genre of computer-mediated communication which is believed to possess a socially-transformative, democratizing potential. Journalists see blogs as alternative sources of news and public opinion. Educators and business people see them as environments for knowledge sharing. Private individuals consider blogs as a vehicle for self-expression and self-empowerment. All of this is purportedly brought about by the technical ability that blogging software affords to update web pages rapidly and easily.

Herring et al.’s (2005) research shows that blogs allow authors to self-express publicly and at the same time experience social interaction without losing control over the communication space. Blogs share lots of features with personal homepages and asynchronous discussion forums, a text-based form of interactive CMC.

Due to the fact that blogs are the latest genre of Internet communication, not many studies can be found in existing literature. Their characteristics are not systematically described either.

### 2.6 Summary

From the above description we get to know that English in computer-mediated environments cover a wide range of language use in our daily life. Each of these genres or situations has its own linguistic features. Of course, they are not the only cases of English in computer-mediated environments. In fact, there are some more. All they have in common is that they are important components of our daily language use. Just like English used in conventional media or situations, English in computer-mediated environments also deserves serious attention from language researchers.

## 3. Representation of computer-mediated English in large corpora

With the advancement of computer technology, corpus linguistics has revived as an important trend in present-day linguistic investigations. Nevertheless, its focus is still on the description and analysis of orthodox texts (e.g., transcriptions of spoken dialogues, books, magazines and newspapers) in their electronic form. In order to find out the status of representation of computer-mediated English in large corpora, the author has checked the composition of five latest leading English corpora and found that most of them have not included any computer-mediated components. In other words, despite the advent of the World Wide Web and CMC since the early 1990s, English in computer-mediated environments has not yet made its way into the focus of linguistic studies. The following tables show the compositions of the currently best-known English corpora in the world.

Table 1 shows the composition of the British National Corpus. The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of current British English, both spoken and written (http://www.natcorp.ox.ac.uk/).

| Medium | Texts | Proportion |
|---|---|---|
| Book | 1488 | 46.36% |
| Periodical | 1167 | 36.36% |
| Misc. unpublished (including email: Leeds United email list 239126 words) | 245 | 7.63% |
| Misc. published | 181 | 5.64% |
| Unclassified | 79 | 2.46% |
| To-be-spoken | 49 | 1.52% |

**Table. 1 Composition of British National Corpus**

In spite of its ambition to represent a wide cross-section of current British English, it only includes a very small part of English in computer-mediated environments: an approximately 240,000-word email component. All the email exchanges are from the Leeds United email list and they are about football. No other subject-matter has been included.

Table 2 displays the composition of the Bank of English, another very influential English corpus which contained 450 million words at its latest count. This corpus was developed primarily for dictionary compilation purposes by Harper-Collins at Birmingham. From the latest dictionaries published by Harper-Collins we can see traces of English in computer-mediated environments. Unfortunately we can see no such traces from Table 2 which is adapted from Krishnamurthy (2002), a consultant of the Bank of English corpus. Chances are more computer-mediated English components have been included in the corpus just as they planned in 2002.

| **SUBCORPUS** | | | **2001** |
|---|---|---|---|
| American Books | 32.44m | 7.23% | 1990 > |
| American Radio (NPR) | 22.23m | 4.96% | 1990-3 |
| BBC World Service | 18.60m | 4.15% | 1990-1 |
| British Books | 43.37m | 9.67% | 1990 > |
| British Ephemera | 4.64m | 1.03% | 1991-6 |
| British Magazines | 44.15m | 9.84% | 1992-00 |
| British Spoken | 20.08m | 4.48% | 1991-6 |
| Economist | 15.72m | 3.50% | 1991-9 |
| Independent | 28.08m | 6.26% | 1995-9 |

| | | | |
|---|---|---|---|
| Times | 51.88m | 11.57% | 1997-01 |
| Guardian | 32.27m | 7.20% | 1995-9 |
| New Scientist | 7.89m | 1.76% | 1992-9 |
| Australian Newspapers | 34.94m | 7.79% | 1995-9 |
| American Ephemera | 3.51m | 0.78% | 1995-6 |
| American Newspapers | 10.00m | 2.23% | 1994-6 |
| Sun and News of the World | 44.76m | 9.98% | 1997-01 |
| American Academic Textbooks | 6.34m | 1.41% | 1990-6 |
| American Spoken | 2.02m | 0.45% | 1994-7 |
| Strathy Canadian Corpus | 15.92m | 3.55% | 1980-00 |
| Wolverhampton Business Corpus | 9.65m | 2.15% | 1999-00 |

**Table. 2 Composition of Bank of English**

Table 3 shows the components of International Corpus of English which was developed in the 1990s. It is a comparable corpus of World English. So far 18 countries from the so-called Inner Circle, Outer Circle, and Expanding Circle are represented. Each regional variety is represented by a million-word subcorpus of both spoken and written language in roughly equal quantities. As can been observed from the table, no components of English in computer-mediated environments have been included.

| | | | |
|---|---|---|---|
| **Spoken** (300) | **Dialogues** (180) | **Private** (100) | Conversations (90)<br>Phone calls (10) |
| | | **Public** (80) | Class Lessons (20)<br>Broadcast Discussions (20)<br>Broadcast Interviews (10)<br>Parliamentary Debates (10)<br>Cross-examinations (10)<br>Business Transactions (10) |
| | **Monologues** (120) | **Unscripted** (70) | Commentaries (20)<br>Unscripted Speeches (30)<br>Demonstrations (10)<br>Legal Presentations (10) |
| | | **Scripted** (50) | Broadcast News (20)<br>Broadcast Talks (20)<br>Non-broadcast Talks (10) |
| **Written** (200) | **Non-printed** (50) | **Student Writing** (20) | Student Essays (10)<br>Exam Scripts (10) |
| | | **Letters** (30) | Social Letters (15)<br>Business Letters (15) |
| | **Printed** (150) | **Academic** (40) | Humanities (10)<br>Social Sciences (10)<br>Natural Sciences (10)<br>Technology (10) |
| | | **Popular** (40) | Humanities (10)<br>Social Sciences (10)<br>Natural Sciences (10)<br>Technology (10) |
| | | **Reportage** (20) | Press reports (20) |
| | | **Instructional** (20) | Administrative Writing (10)<br>Skills/hobbies (10) |
| | | **Persuasive** (10) | Editorials (10) |
| | | **Creative** (20) | Novels (20) |
| *****(Numbers in brackets indicate the number of 2,000-word texts in each category). | | | |

**Table. 3 Composition of International Corpus of English**

Table 4 shows the structure of Cambridge International Corpus built up by Cambridge University Press over the last ten years to help writing books for learners of English. The English in this corpus comes from newspapers, best-selling novels, non-fiction books on a wide range of topics, websites, magazines, junk mail, TV and radio programmes, recordings of people's everyday conversations and many other sources. The corpus has over 700 million words and it will continue to grow each year as new data is added. Websites have become one of the data sources, but no further information about what web contents have been included can be obtained.

| **British English** | |
|---|---|
| **No. of words** | **Corpus** |
| 450 million | Written British English |
| 17 million | Spoken British English including the unique CANCODE corpus, collected jointly by Cambridge University Press and the University of Nottingham |
| 20 million | Written British academic English |
| 30 million | Written British business English |
| 1 million | Spoken British business English – CANBEC – The Cambridge and Nottingham spoken Business English Corpus |
| **American English** | |
| **No. of words** | **Corpus** |
| 200 million | Written American English |
| 22 million | Spoken American English including the Cambridge-Cornell Corpus of Spoken North American English collected jointly by Cambridge University Press and Cornell University in the United States |
| 7 million | Written American academic English |
| 30 million | Written American business English |
| **Learner English** | |
| **No. of words** | **Corpus** |
| 19 million | Learners' written English (the Cambridge Learner Corpus) |
| 8 million | Error coded learner written English |

**Table. 4 Composition of Cambridge International Corpus**

From the above four large English corpora we get to know that English in computer-mediated environments has not attracted adequate attention of large English corpus compilers, though there seem to be signs of selectively including data from websites.

## 4. The notion of representativeness revisited

English in computer-mediated environments poses lots of questions for corpus linguistics. One of these questions is how to understand the notion of representativeness in an ever-changing linguistic environment. For a general-purpose corpus, can we still say it is representative if it does not include any component of computer-mediated English at all? The answer is negative. According to Biber (1993) when he is discussing the shortcomings of proportional language corpora, corpus for linguistic research requires language samples that are representative in the sense that *they include the full range of linguistic variation existing in a language*. Obviously all the corpora mentioned above failed to meet this requirement properly. One main reason for that would be these corpora were started five or even ten years ago. At that time, English in computer-mediated environments was less common and less influential than nowadays. Now, computer-mediated communication has become an integral part of many people’s daily life. It is necessary to take English in computer-mediated environments into consideration if we want to make English corpora more representative.

## 5. Necessity of constructing an international corpus of computer-mediated English

The advent of the Internet, the popularity of computer-mediated communication, and the globalization of international economy have begun to blur the boundaries among the major regional varieties of English. English in computer-mediated environments is no longer a language which only belongs to people from the UK, the US, Canada, Australia, and New Zealand; it has become an international language. This language has not only inherited the core features of all those major regional varieties but also taken on many new features from various sources. As Ide, et al. (2002) have admitted, it is very difficult to differentiate one variety of English from another in a computer-mediated environments. English data obtained from American websites or US-based chatrooms may not necessarily be American English. One question we should ask ourselves is that how necessary it is for us to talk about concepts such as regional varieties in a community where the boundaries between nations and peoples are blurred? Are we actually having a new variety, so to speak, which is different from all the established regional varieties? To answer this question, we need to do plenty of research. One way of doing it is to construct an international corpus of English used in computer-mediated environments and carry out systematic investigation about the linguistic features of this kind of English and its subvarieties. This corpus can be either used as a supplement to the International Corpus of English or used on its own for investigating common linguistic features of international English used in computer-mediated environments. It can also be used as a data source for making dictionaries of International English and a resource for teaching English as an international language.

## 6. Problems with building such a corpus

As Kilgarriff and Grefenstette pointed out, the Web is immense, free, and available by a mouse click. It contains hundreds of billions of words of text and can be used for all manner of language research (2003: 333). Nevertheless, a lot of important decisions must be made before we can start constructing a corpus out of it. For example, we need to decide on the type of corpus we want to create: a conventional corpus or a multi-modal corpus? If it is the former, we then need to decide on which subvarieties to be included, how much data for each subvariety is adequate, and how big the total size of the corpus would be. If it is the latter, we then need to decide how to record and represent features other than texts, in what format? Whether the corpus will be annotated? If yes, what to annotate and how? Besides, we need to settle copyright problems, fund problems, and many other problems.

## 7. Conclusion

The advent of the Internet and computer-mediated communication has considerably changed the linguistic environment we are in. As a result, English in computer-mediated environments should no longer be ignored by researchers in linguistic studies. In order to properly describe the English language as a whole, we should either incorporate computer-mediated English components into our general-purpose corpus construction or build an international corpus of English in computer-mediated environments so that we can carry out detailed and systematic research about the newly emergent linguistic dimension.

This paper raises more questions than offer solutions. The main purpose of so doing is to raise our awareness of the altered linguistic reality so that we can adjust our practice in corpus linguistics and linguistic studies accordingly.

## References

Biber, D. (1993) Representativeness in corpus design. *Literacy & Linguistic Computing*, 8(4), 243-257.

Crystal, D. (2001) *Language and the Internet* (Cambridge: Cambridge University Press).

Crystal, D. (2003) *The Cambridge Encyclopaedia of the English Language* (Cambridge: Cambridge University Press).

Herring, S. C. (Ed.). (1996) *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives* (Amsterdam: John Benjamins).

Herring, S. C. (1999). Interactional Coherence in CMC. *Journal of Computer-mediated Communication*, 4 (4)

Herring, S. C. (2001). Computer-mediated discourse. In: D. Schiffrin, D. Tannen, and H. Hamilton (Eds.), *The Handbook of Discourse Analysis* (Oxford: Blackwell Publishers), 612-634.

Herring, S. C., Scheidt, L. A., Bonus, S., and Wright, E. (2005) Weblogs as a bridging genre. *Information, Technology & People*.

Hutchby, Ian. (2001) *Conversation and technology: from the telephone to the Internet* (Blackwell Publishers).

Ide, N., Reppen, R., Suderman, K. (2002) The American National Corpus: More Than the Web Can Provide. *Proceedings of the Third Language Resources and Evaluation Conference* (LREC), Las Palmas, Canary Islands, Spain, 839-44. Available on-line from http://www.cs.vassar.edu/faculty/ide/pubs.html (accessed June 10, 2005)

Kilgirraff, A. and Grefenstette, G. (2003) Introduction to the Special Issue on the Web as Corpus. *Computational Linguistics* 29 (3), 333-347.

Krishnamurthy, R. (2002) The Bank of English past, present, and future: corpus size, composition, annotation and software. Available online from http://www.dcs.shef.ac.uk/research/ilash/Seminars/rameshLR2.doc (accessed June 13, 2005)

Macfadyen, L. P. and Doff, S. (2003) The Language of Cyberspace: Text, Discourse, Cultural Tool. Available online from http://station03.olt.ubc.ca/index.php?title=The_Language_of_Cyberspace (accessed June 14, 2005)

Ooi, V. B. Y. (2002) Aspects of Computer-mediated Communication for Research in Corpus Linguistics, in Peters P, P Collins & A Smith (eds.) *New Frontiers of Corpus Research* (Amsterdam-New York: Rodopi), 91-104.

Schegloff, E. A. (1968). Sequencing in Conversational Openings. *American Anthropologist* 70, 1075-95.