

“Synchronous online chat” English: Computer-mediated communication

RAMI A. AL-SA'DI* and JIHAD M. HAMDAN**

ABSTRACT: This study aims to investigate the linguistic features of e-English, relying on a large corpus of authentic IRC and Yahoo Messenger chat sessions (both public and private). The study concludes that economy is a salient attribute of e-English. Among the other major linguistic features that were detected in this study are: sentences are characteristically short and simple, many words are distorted and truncated in familiar and unfamiliar ways, abbreviations and acronyms are widespread, and taboo words very likely to occur in most chat sessions (especially in public ones). One more major finding is that e-English can be viewed as unclassifiable as either written or spoken language despite the written medium wherein it occurs.

INTRODUCTION

This paper describes the main linguistic features of a new mode of English, namely the English used in Internet real-time chat channels/rooms.¹ In particular, it focuses on two manifestations of synchronous Computer-Mediated Communication (CMC): Yahoo Messenger and Internet Relay Chat (henceforth Y! and IRC, respectively).

The dramatic evolution in Cyberspace has left its impact on language at all levels: syntactic, semantic, morphological, and most importantly lexical. Halle (n.d.) in his comment on the role of the Internet in language change, suggests that the Internet is a distinct language domain with its own linguistic features and advantages.

To exemplify the lexis of Cyber-English, one may wish to look into the features of *pic* and its two possible plurals *pics* and *pix*. For instance, in the 1974 edition of the *Oxford Advanced Learner's Dictionary (OALD)* none of these three items is recognized as an English word. In the 1995 edition of *OALD*, the word *pic* has entered the dictionary as an informal synonym for *picture*, but it is still null of a plural form. The 1995 edition of the *Longman Dictionary of Contemporary English (LDCE)* has an entry for *pix* only as a colloquial plural word meaning *pictures* or *photographs*, but it lacks a singular form.

The 1993 edition of *Webster's Dictionary and Thesaurus (WDT)* states that *pixel*, which is any of a number of very small picture elements that make up a picture, as on a visual display unit, has its origin in *pix* (pictures) and *elements*. Therefore, *WDT* is a pioneer in clearly denoting that technology is the origin of *pix* and also *pixel*. The three forms, *pic* and its two plural forms, are recognized by the 2000 edition of *Collins English Dictionary (CED)* as English words that entered the language in the late twentieth century. This word, however, is indigenous to Cyber-English, and the use of *picture*, *photo*, or *photograph* might in some instances be an indicator of inexperience with e-chat.

Being a means of synchronous communication, online chat may logically be perceived as a simulator of Real-Life (RL), face-to-face (FTF) conversation. This deduction which

* Faculty of Arts, University of Jordan, Amman, Jordan. E-mail: enigma_1g@yahoo.co.uk

** Faculty of Arts, University of Jordan, Amman, Jordan. E-mail: jihadan@hotmail.com

we can cherish for the time being implies that online chat English resembles spoken English to a great extent. When using IRC or Y!, one types a message which would be received by one's interlocutor(s) as soon as one presses the "Enter" button on the keyboard, and it is natural to expect an immediate reply. This process resembles turn-taking in RL conversations.

Another proposition supporting the speech-like nature of online chat English is that there is always room for "a slip of the tongue." In online chat, one might "say" something one does not mean to say, and accordingly either apologize, rephrase one's statements, or try to thwart any impending misunderstanding by utilizing any possible means at one's disposal. Apparently, astute talkers in RL, FTF conversation are more likely to be astute "talkers" in synchronous CMC.

Rinvoluceri (n.d.) demonstrates that a salient feature of spontaneous FTF spoken English is "the opportunity for repair and paraphrase," employed auspiciously and instantaneously to avoid misinterpretation by the listener and so to avoid censure. The listener's reactions and feedback do have immense impact on the shaping up of the text-producer's wording, setting forth of ideas, speed of enunciating words and phrases, etc. In the written medium, however, there is no such immediate opportunity for the sender to monitor the reader's feedback and effect repair. Clearly, the affinities between spontaneous spoken English and synchronous Internet chat English surpass one's expectations, for what has been discussed is merely a sketch of the similarity between spoken English in a verbal interaction and synchronous CMC English for Cyber-chat purposes.

OBJECTIVES AND SIGNIFICANCE

This study investigates the main linguistic features of authentic IRC and Y! English chats, both public and private. In particular, it focuses on sentence structure, i.e. sentence length and sentence complexity, in addition to lexical features, i.e. word truncation, orthography, word formation and taboo words. It also attempts, in the course of discussion, to examine the similarities and differences between Cyber-English on the one hand and each of written and spoken English on the other.

E-chat as a CMC form of English is spreading at such record speed that it has become worthy of in-depth study. That IRC and Y! English has not yet been standardized warrants further changes to come and suggests that some (or maybe many) of its current features are ephemeral and are consequently doomed to collapse and be replaced by others in due course. Hence, the present linguistic features of this special type of online chat English will be highlighted. To the researchers' knowledge, the study reported here is the first of its type that solely addresses itself to this task.

CORPUS

The study was not intended to be carried out on a certain number or group of IRCers and Y! users, but rather on a certain number of authentic chats both private and public. This seemed necessary to ensure diversity of users and interests and to avoid emphasis on a particular group of chatters with a limited set of topics and expectations. Hence, the corpus comprised 20 actual chats: 10 private and 10 public (see Appendices I and II for a private chat sample and a truncated public chat sample). A major reason for the inclusion of both types is that the language of private chats may differ significantly from that of

public chats. Keeping in mind language and gender issues (Kaplan, 1989), the researchers used both “bisexual” and “unisexual” samples in the private chat category.² As for the public ones, mixed-sex channels and rooms seem to be the norm.

In private chats, the Y! IDs and IRC nicknames of the participants were excised and replaced by symbols at those participants’ request. Public chats, however, give all the IDs and nicknames of the participants.

Since one of the researchers had to use his own Y! ID and to use an IRC nickname so as to be able to log in to public chat channels, he only lurked there without taking part in any discussion and ignored any postings he received from e-chatters while the recording of the session was going on. That had to be done to avoid incorporating this researcher’s own language into the corpus. The private chats were provided by the chatters themselves who, for one reason or another, had already recorded their chats. This arrangement was based on confidence and assurance that the chats would be used for research purposes only. The actual chats used as a corpus of the study are as follows:

- Private chats: 10 samples from both IRC and Y!, ranging in length from 2 to 15 pages each.
- Public chats: 10 samples, halved among IRC and Y!. These samples range in length from 10 to 30 pages each.³

RESULTS AND DISCUSSION

The linguistic features that are going to be explored include sentence structure (sentence length and sentence complexity) and lexical features (word truncation, orthography, word formation and taboo words). Moreover, a special section provides an overview of the similarities and differences between Cyber-English and written English on the one hand and Cyber-English and spoken English on the other.

Sentence structure

Sentences will be examined in this section for their length (average number of words in each sentence) and syntactic complexity (simple, compound, complex, or compound-complex).

Sentence length. Before categorizing sentences as either long or short, two questions present themselves: How many words should a sentence have in order to be considered long or short? What determines sentence boundary in e-English? Werry (2003) puts the number of words for a “long” sentence at six or more. In light of this, the following sentence will be viewed as long:

- (1) i was doin jus the same after that sweet chat

Nonetheless, many people may not regard this sentence as a long one, even though it consists of ten words.⁴ It may be viewed as a sentence of medium, ordinary length. This brief argument relies almost entirely on sheer speculation; however, it may lead one to conclude that a six-word sentence is too short, a ten-word sentence not at all long. Therefore, the researchers have, somehow arbitrarily, adopted the number eight for words as the dividing boundary between “short” and “non-short” sentences. For convenience, non-short sentences will be termed “long sentences” in this work.

The answer to the second question above does not seem to pose a very serious problem. Conventionally, a sentence in writing (and the form of English under investigation is all written) is any well-formed and meaningful string of words which starts with an upper-case letter and ends with a full stop. Of course, a one-word piece of language containing a gap-filler, a brief response to a statement or question, or an interjection (yes, well, really, aha, etc.) will be regarded as a full sentence in this piece of research. Nonetheless, this compressed solution does not come to grips with the task of determining sentence boundary in e-English. This may be due to the absence of punctuation marks in much of Cyber-English or to the use of improper punctuation and spelling. The following example illustrates how defective (or inadequate) punctuation hinders dividing e-posts into sentences:

- (2) <fransiska>lol since u brought up the name of bin laden i must say it is no entirely his parents fault.. lol he is one of 52 siblings!

The exact function of the five dots that follow the acronym "lol" (laughing out loud) is not easy to tell. They could designate a pause or, less likely, hesitation, but they do not seem to suggest an ellipped element, and after all in written discourse only three spaced dots (called ellipsis dots), and not five, are used for this purpose (Heffernan and Lincoln, 1990: 460). Hence, this acronym may be viewed as a separate item that compensates for the lack of paralinguistic cues, as it is reasonable to assume that by using "lol," <fransiska> is telling us: I'm laughing while saying this!

Moreover, the final portion of <fransiska>'s post – *he is one of 52 siblings!* – is an independent clause and can be a sentence on its own, but it can also be a part of a complex sentence since we can view the post as consisting of two subordinate clauses, i.e., *since u brought up the name of bin laden* and [because] *lol he is one of 52 siblings*, and one independent clause, i.e., *i must say it is no entirely his parents fault*.

In most posts punctuation is missing, which makes unit breaks difficult to recognize. Therefore, it is quite rational to employ the most accepted linguistic norms and reasoning in trying to divide stretches of words into sentences. Admittedly, this plan does not always work well, but it appears to be the least risky approach for the purpose of this piece of research.

Another relevant point is that a few posts can never be conceived of as full sentences or clauses, nor are they short answers to previously posted questions, as in (3):

- (3) sorry_name_taken_already: anyone from

Other posts are totally incomprehensible (or are not worthy of taking into account), since they are either incomplete sentences, over-misspelt, or they are not English in the first place:

- (4) a. drake: I woud bo fa ak sty
b. JUNKIE: ad eveddrpo wen da

Inevitably, there is no way to consider such posts as sentences since they defy comprehension or are non-English. Such posts were excluded from sentence analysis.

The corpus contained 5,591 sentences, 4,829 of them (86%) were short and 762 (14%) long. Tables 1 and 2 show the number and percentage of short and long sentences in the corpus. Tables 1 and 2 make it evident that short sentences are the overwhelming majority in the corpus as a whole and in each sample as well. In private samples, 86% of the sentences are short, 14% long. In public samples, 87% are short, 13% long.

Table 1. Number and percentage of short and long sentences in private chats

Chat type and number	No. of sentences	Short sentences		Long sentences		No. of non-sentences
		No.	%	No.	%	
IRC						
1	107	92	86	15	14	11
2	52	45	87	7	13	5
3	145	114	79	31	21	26
Y!						
1	452	407	90	45	10	13
2	33	25	76	8	24	1
3	114	84	74	30	26	14
4	154	111	72	43	28	0
5	198	176	89	22	11	2
6	184	178	97	6	3	0
7	148	127	86	21	14	11
Total	1,587	1,359	86	228	14	83

A straightforward explanation of this finding is as follows. It has been noted that users of IRC and Y! send their typed posts in a chatter that looks very much like FTF conversation (see Introduction). Typing speed, even for touch-typists, can never match the speed of oral enunciation of sounds. One major means for making up for this deterrence is to type as fast as one can. However, whether a Cyberer is able to type fairly fast with a modicum of effort or finds speedy typing an inaccessible gift, it should be natural to expect many (and probably most) e-chatters to resort to short, succinct sentences to save time and effort.

Sentence complexity. Before looking at the distribution of sentences in terms of structure complexity, it should be noted that: Firstly, English sentences in this regard

Table 2. Number and percentage of short and long sentences in public chats

Channel/Chat room	No. of sentences	Short sentences		Long sentences		No. of non-sentences
		No.	%	No.	%	
IRC						
London	368	332	90	36	10	81
Cybersex	439	411	94	28	6	81
Chatzone	165	155	94	10	6	23
Chataway	326	247	76	79	24	24
Britney Spears	601	520	87	81	13	100
Y!						
Professors' Chat Room	677	550	81	127	19	108
Islam Chat	518	420	81	98	19	54
20's Love Room	505	452	90	53	10	32
Lesbian Women's Haven	233	222	95	11	5	13
Australia	176	161	91	15	9	25
Total	4,008	3,470	87	538	13	541

are divided into four types, i.e., simple (containing just one clause), compound (containing two conjoined clauses), complex (containing at least one embedded clause with a grammatical function), compound-complex. Secondly, in spoken English, which Cyber-English presumably has much in common with, compound sentences enable better comprehensibility than do complex ones, and therefore the former type outnumbers the latter (Quirk *et al.*, 1972: 795). So, initially, it can be presupposed that simple and compound sentences in e-English are the vast majority, whereas the other two types, which often require more effort and time in processing, typing and reading, are naturally meagre. It is true that a complex sentence may in some cases be shorter and less effort- and time-consuming than a simple or compound sentence, but the above argument has only been introduced in general terms. Thirdly, some sentences in the corpus, as is the case in spoken English and many types of written English, consist of only a subordinate clause and yet they are not perceived as fragmentary or ungrammatical. That is because they may be replies to previously posed statements or questions, and they imply a sense of ellipsis.

- (5) <A> y r u on invisible?⁵
 cuz I don't wana talk to my ex

's sentence is understood to have an ellipted element: (*I'm on invisible*) *cuz I don't wana talk to my ex*. Therefore, the best way out is to categorize this sentence as a complex sentence despite the absence of an independent clause because it is recoverable.

Tables 3 and 4 show the distribution of sentences in terms of structure complexity. The final column was incorporated into the tables to make collating simple sentences with the other sentence types easier. The tables show that simple sentences constitute the overwhelming majority of the corpus; their percentages in private and public chats are 92% and 93%, respectively. The difference between the frequency of occurrence of compound sentences and complex ones is minimal and rather negligible whether the chat is private or public. The percentage of complex sentences in private chats exceeds that of compound sentences by 1%, and in public chats by 2%. This finding, however, should not lead us to

Table 3. Number and percentage of sentences according to complexity in private chats

Sample chat	No. of sentences	Simple (1)		Compound (2)		Complex (3)		Compound-complex (4)		Non-simple sentences 2 + 3 + 4	
		No.	%	No.	%	No.	%	No.	%	No.	%
IRC											
1	107	96	90	3	3.0	7	7	1	1.00	11	10
2	52	47	90	5	10.0	0	0	0	0.00	5	10
3	145	123	85	13	9.0	9	6	0	0.00	22	15
Y!											
1	452	436	96	2	0.4	14	3	0	0.00	16	4
2	33	29	88	1	3.0	3	9	0	0.00	4	12
3	114	96	84	10	9.0	5	4	3	3.00	18	16.
4	154	134	87	9	6.0	9	6	2	1.00	20	13
5	198	187	94	3	2.0	6	3	2	1.00	11	6
6	184	181	98	0	0.0	3	2	0	0.00	3	2
7	148	132	89	8	5.0	7	5	1	0.70	16	11
Total	1587	1461	92	54	3.0	63	4	9	0.06	126	8

Table 4. Number and percentage of sentences according to complexity in public samples

Sample chat & room	No. of sentences	Simple (1)		Compound (2)		Complex (3)		Compound-complex (4)		Non-simple sentences 2 + 3 + 4	
		No.	%	No.	%	No.	%	No.	%	No.	%
IRC											
London	368	347	94.0	5	2.0	16	4.00	0	0.0	21	5.71
Cybersex	439	414	94.3	3	0.7	20	4.60	2	0.5	25	5.70
Chatzone	165	161	97.6	1	0.6	3	1.80	0	0.0	4	2.40
Chataway	326	300	92.0	6	1.8	16	4.90	4	1.2	26	8.00
Britney Spears	601	559	93.0	14	2.3	26	4.30	2	0.3	42	7.00
Y!											
Professors' Chat Room	677	595	87.9	33	4.9	40	5.90	9	1.3	82	12.10
Islam Chat	518	481	92.9	11	2.1	23	4.40	3	0.6	37	7.10
20's Love Room	505	474	93.9	11	2.2	18	3.60	2	0.4	31	6.10
Lesbian	233	226	97.0	3	1.3	4	1.70	0	0.0	7	3.00
Women's Haven											
Australia	176	169	96.0	3	1.7	4	2.30	0	0.0	7	4.00
Total	4008	3726	93.0	90	2.6	170	4.24	22	0.6	282	7.00

dismiss the idea that in both spoken English and Cyber-English, compound sentences may naturally outnumber complex ones. In this regard, the researchers wish to report that they decided to envision each two independent clauses, unless they are linked by a coordinator or properly punctuated, as two simple sentences rather than one compound sentence. Another independent look at the data may suggest a slightly different conclusion. Finally, compound-complex sentences are so rare; their percentages are 0.57% and 0.55% for private and public chats, respectively.

The ultimate conclusion is that the simple sentence is the prevailing sentence type in e-English, for the percentages of simple sentences for both private and public samples are 92% and 93%, respectively. Complex and compound sentences occur equally frequently as long as we cherish the idea that Internet-punctuation (or lack of it) sometimes tempts us to strip down each potential compound sentence to several simple sentences; and compound-complex sentences occur only so sparingly that they are made negligible.

Lexical features

This section examines the various methods of word truncation, the idiosyncrasies of Cyber-orthography, the processes of word formation in Cyberspace, and the prevalence and causes of the use of taboo words.

Word truncation. A close examination of the data showed that Cyber-chatters tended to utilize manifold methods of word truncation in order to lessen the number of keystrokes and to keep the flow of the chat. Nonetheless, Cyberers are not all equally keen on truncating words. An e-chatter who tends to use the full forms of words is called a newbie (Crystal, 2001). Newbies are usually more adept at writing “natural” English than

at using Cyber-orthography. Some of them reported (personal communication) that when they encounter strangely spelt words posted by old-timers, they may be shocked and fail to grasp the real word intended, e.g.:

- (6) <fransiska> skeppy.. i think u r pickin a fight just to let go of some aggresion or issues u have in quite another context.wana take it out on me? Well, I am strong enough. Do ur worst.
<alman4ever> Fransiska, you cannot spell right.

However, some e-chatters, though not newbies, might choose to spell their words correctly and never shorten them, to make sure that they are making themselves understood by everyone, to demonstrate that they are much faster typists than the others, to gain the admiration of the others for using “good” English, etc. Below is an example from an old-timer:

- (7) shawtboi: I'm an American at Harvard, and I major in Biology and Chemistry, and with the terms we use i would not say America has a dense vocabulary but only common words are understood in society.

Generally, the norm remains for words to be shortened whenever possible for most e-chatters. Following are the methods of word truncation that we were able to identify.

1. using one letter or a numeral *in lieu* of a word when the letter or numeral and the word are homophones in speech:
 - (8) a. <A> ok y r my pms on? [*y r* for *why are*]
 - b. WHAT R U DOIN? [*r u* for *are you*]
2. replacing a syllable in a polysyllabic word by a homophonous letter or numeral:
 - (9) a. holydevil1981: 2morrow will come [2 stands for “to . . .” in *tomorrow*]
 - b. u should've told me 15 mins b4 [*b* stands for “be-,” 4 for “-fore” in *before*]
3. deleting all (or most) of the vowels in a word as long as the consonants are still capable of unambiguously conveying the intended meaning:
 - (10) a. <A> to play bsketbll with his baseball team [*bsketbll* for *basketball*]
 - b. <A> all the normal ones i have.. have other ppl in em [*ppl* for *people*; *em* for *them*]
4. using one letter for a word (usually a function word rather than a lexical one):
 - (11) <anthrax* e-mailer> i start off long n slow than go short n fast [*n* for *and*]
5. g-dropping in -ing constructions:
 - (12) <amyfoxxpretty> u ignore the sufferin of the hungry [*sufferin* for *suffering*]
6. using informal, colloquial, or slang spelling if shorter than the standard spelling of the same word:
 - (13) a. oh cuz it's vday and where'd he tell me he was going [*cuz* for *because*]
 - b. I JUST WANNA KNOW [*wanna* for *want to*]
7. using only the initial letter of a word when the context easily determines the intended word:
 - (14) IN 2WEEKS ITS MY BDAY I WANNA GO OUT [bday: birthday]
8. spelling a word just as it is pronounced, making Cyber-English spelling a good guide to pronunciation (vis-à-vis Standard English):
 - (15) a. LATA [*lata* for *later*]
 - b. U REMEMBER OUR LIL DEBATES? [*lil* for *little*]

(16) <hulette> wots ur country? [*wots* for *What’s*]

Cyber-orthography. This section is strongly linked to the previous one, however, it is rather more dedicated to spelling and other orthographic conventions. Below are the major Cyber-spelling conventions:

1. Capitalization is rare. Even proper nouns and sentence-initial words are usually typed with a lower-case initial letter:

(17) it’s funny. i think christian has got the whole game under hi belt

2. Sometimes a whole sentence (or a series of sentences) is typed in capitals. That is because the user has his “caps on.” It could be meaningless, but it is sometimes used to show emphasis or to denote shouting, anger, etc.:

(18) A: U SHOULD HAVE DONE THAT FROM BEGINNING

B: when one use capital letters in chatting it means he or she is shouting from anger

3. Contracted forms often lack an apostrophe:

(19) <neanderthals_need_not_apply> n twice as good now that ive over-explained it, eh?

4. Non-linguistic human sounds that are usually described in writing (e.g., in drama) are simulated with English characters in Cyber-English. Contrast (20a) and (20b); the first taken from a play, and the other from a Cyber-sample:

(20) a. Miss Y. (Shrieks with laughter.)

b. <the_real_female_slim_shady> hahahahahahaha

Word formation in Cyberspace. Cyberers tend to use regular word formation processes in English to create their own Cyber words. Below is a brief presentation and discussion of these processes.

1. Coinage. One example is “iggy.” This word entered e-English as a noun meaning *blockhead*, *moron*, etc., and was mainly used to refer to newbies who display maudlin behaviour as to their incapability of using the software.

2. Broadening or extension. Gradually, *iggy* started to give ground to other meanings too as it underwent broadening (see Freeborn, 1994 for a description of semantic broadening of English vocabulary). After the broadening process, *iggy* came to be an unfriendly word which can be used in “flaming” against any Cyberer, be he a newbie or a highly experienced e-chatter:

(21) <steal_ur_hart> porno girls r automated programmes, and anyone who pays them attention is a iggy

An interesting point not to be missed here is that while diachronic changes in the semantics of words take place over several years, sometimes centuries, semantic (as well as other linguistic) changes in Cyber-English may occur within a matter of months. Another semantic broadening *iggy* underwent is its association with “ignoring.” Therefore, to block someone’s private messages so as not to receive any of them, all one has to do is “put him on the ignore” or “put him on iggy,” or, put differently, a repulsive e-chatter may “get an iggy.”

3. Conversion. The same Cyber-word discussed above is a case in point. No sooner had *iggy* come into Cyber-being than e-chatters started using it as a verb too. Accordingly, unpleasant e-chatters can be “iggied” by others. One more interesting fact about *iggy* is that it still awaits to be incorporated into dictionaries. For instance, it is not recognized by the 2000 editions of both *OALD* and *The American Heritage Dictionary of the English Language*.

4. Backformation followed by conversion. An interesting example is that of *Cyber*, which came as a backformation from *Cybernetics*. Then *cyber* was converted into different open classes:
 - (22) a. dhriti2003: looking for a dream cyber guy [Adjective]
 - b. lollypop: 5 minutes left. who wnts to cyber? [Verb]
5. Derivation. For instance, *Cyber* can receive the agentive suffix to produce *Cyberer*, used in this paper as a synonym for "e-chatter." The same suffix was attached to IRC (IRCer) to refer to users of this CMC medium.
6. Clipping:
 - (23) cyrustemplar: does amy talk with her mouth full? [*amy* for the nickname *amyfoxxypretty*]
7. Acronyms:
 - (24) a. lol which game r u referin to? [lol: laughing out loud]
 - b. brb [brb: be right back]

In fact, some of these abbreviations, and acronyms are indigenous to Cyberspace and are still confined to it.

As is clear, Cyberspace is a fertile environment for the emergence of words. However, dictionaries, especially paper-based ones, have not admitted many of them. This may be ascribed to the fact that these words are still brand new and are used mainly by a certain group of people, and hence cannot be straightforwardly adopted as "new" English words. Another reason for the slow entry of Cyber-words into dictionaries is the "fleeting text" where they were born. As soon as a chat session is over, it all disappears unless it is recorded.

Taboo words. One might ask: What assigns *tabooness* to words? Following are the entries for *taboo* in some dictionaries, not just for the purpose of dictionary definition but to identify any significant conditions that a word must fulfill to be a full-fledged taboo.⁶

LDCE (online): 1a subject/area/word etc. that people avoid because they think it is offensive or embarrassing. 2 not accepted as socially correct. 3 too holy or evil to be touched or used.

CEC (2000): 1 forbidden or disapproved of. . . 3 any prohibition resulting from social or other conventions. 4 ritual restriction or prohibition, esp of something considered holy or unclean.

OALD (2000): 1 a cultural or religious custom that does not allow people to do, use or talk about a particular thing as people find it offensive or embarrassing.

taboo words words that many people consider offensive or shocking, for example because they refer to sex, the body or people's race

Some definitions show that taboo words are those that cause offensiveness or embarrassment. But are taboo words offensive and embarrassing in all contexts for all people alike? Certainly not, since taboos are culture-bound, and therefore they vary from culture to culture. Apparently, any attempt to inflict the mores of any social group upon a multi-cultural community, in this case the Cyber-community, would be a fiasco. This implies that in such a setting only universal taboos remain. But "universal taboos" is just another contentious term to struggle with.

The initial question remains unanswered: what words in the corpus of this study are to be considered as taboos? In the light of all the arguments above, it would be useful to regard the following as tentative categories of taboo words/expressions:

1. blasphemous words, i.e., those that profane the name of God or any sacred objects/concepts in any religion, e.g., *chat god*;
2. four-letter words that are recognized by almost everyone as taboo, and any longer compound words to which a four-letter word has been attached, e.g., *ass* and *dumbass*;
3. racially prejudiced words, e.g., *sand nigga*;
4. religiously prejudiced terms, e.g., *muslame*;
5. any other items that many people would regard as “swear words” or very repulsive/impolite words or words labelled as taboo in dictionaries, e.g., *blowjob*.

Tables 5 and 6 show the frequency of occurrence of taboo words in the corpus. For each sample, frequency can be roughly illustrated through the number of taboos in proportion to the total number of sentences. It is sensible to perceive a sample as dreadfully lecherous if it shows, on average, one taboo word per ten or less than ten sentences. A look at Table 5 shows that the use of taboos in e-chat is not a must in all cases. There are four private sample chats that are null of taboo words, whereas Table 6 shows that complete absence of taboos in public chat channels is very improbable. The non-use of taboos (and whenever *taboo* is mentioned in this work, the verbal type is of course meant) in four private chats can be attributed to the belief that since private chats are carried out by only two people, it is often the case that they ostensibly decide and agree on what type of language they should (or prefer) to use. It was initially hypothesized that if at least one taboo word/ expression occurs in each ten or less than ten sentences on average, then the corpus is “dreadfully lecherous”; in reality the corpus confirms this hypothesis as the statistics reveal.

Internet chat English versus written and spoken English

Written and spoken English are language varieties according to medium (Quirk *et al.*, 1972: 22). But is it possible for e-chat English to be classified unequivocally as either a

Table 5. Number of taboo words as well as number of sentences per each one taboo word in private samples

Sample chat	No. of sentences	No. of taboos	No. of sentences per each one taboo
IRC			
1	107	13	8.2
2	52	12	4.3
3	145	25	5.8
Y!			
1	452	85	5.3
2	33	7	4.7
3	114	13	8.8
4	154	0	–
5	198	0	–
6	184	0	–
7	148	0	–
Total	1587	155	10.2

Table 6. Number of taboo words as well as number of sentences per each one taboo word in public samples

Channel/chat room	No. of sentences	No. of taboos	No. of sentences per each one taboo
IRC			
London	368	33	11.2
Cybersex	439	56	7.8
Chatzone	165	12	13.8
Chataway	326	11	29.6
Britney Spears	601	87	6.9
Y!			
Professors' Chat Room	677	2	338.5
Islam Chat	518	93	5.6
20's Love Room	505	37	13.6
Lesbian Women's Haven	233	17	13.7
Australia	176	53	3.3
Total	4008	401	10.0

written or spoken mode or somewhere in between (see Introduction)? Mizrach (2003) touches upon such concerns. One of the similarities between e-mail English and spoken English that he attempts to make explicit is that in e-mail there is immediacy and directness of speech, for one writes not for the public, as in the case of most types of writing,⁷ and not for a relative or a pen-pal who will reply after several days or even months, as in the case of personal letters. However, Mizrach makes it clear that e-mail shares features with spoken English, too. For instance, e-mail "has the deliberative quality of letter writing, in that a person can revise repeatedly what they want to 'say' before 'saying' it."

Werry (2003) investigated the possible motivations that impel CMC chat users to deviate from the norms of the written medium. December (1996) observed that both e-chat and oral discourse are characterized by abbreviation and hence speed. He concluded that on-line chat is unique writing in that it allows for explicit and emphatic sociability and rapidity.

The most obvious characteristic of e-English that would categorize it as a written mode is that it is written in the first place. However, it is written on a computer screen, and it is on computer screens that it is read, and so it differs from traditional written English in that it is not written on paper. Moreover, the written mode normally assumes the addressee to be absent (Quirk *et al.*, 1972). In contrast, the addressee in e-English is only physically "absent," but he must be "present," sitting by his computer screen right at the moment when a message is posted. If the e-addressee is not online, no language is written and no post is sent (except for offline messages, which resemble brief voice messages on the voice mail). This observation sets IRC and Y! English starkly apart from the written mode.

Another aspect of writing is that it is a solitary act. Heffernan and Lincoln (1990: 3) posited that in an FTF conversation, one normally "talk[s] to someone who talks back, who raises questions, who lets you know whether or not you are making yourself clear. But when you write, you work alone." This fact would toss e-English into the spoken

medium rather than otherwise. Therefore, writing on the Net is not a solitary act, but a real-time interaction.

This argument implies that Cyber-English has more in common with spoken English than with written English, but this is not exactly true. For one thing, human voice is lacking. Interlocutors are totally silent, which simply entails loss of many elements that expedite conversation. For instance, paraverbal cues such as gestures, voice quality, accent, and intonation are missing (Dean and Wilson, 1964). Dexterous e-chatters try to make up for this deficit via a plethora of means, e.g., using emoticons (the typographical symbolization of emotion), typing in a certain font style, colour, and size. One may also use italicization, emboldenment, and underlining of (parts of) one's posts to highlight elements or to achieve any other effects. But these means, save for emoticons, make e-English more written-like than spoken-like.

Heffernan and Lincoln (1990) added that writing gives the writer time to think, to revise and edit the output, etc. To some extent, this facet is applicable to Cyber-English. It is true that there is more time available for a Cyberer than there is for a talker in an FTF conversation to think and rethink before giving out words, but there are time limits though. For instance, too much thinking requires too much time; this in turn means very slow typing, which is intolerable for many Cyberers.

Some people tend to view writing as socially superior to speech (Brook, 1974: 146). However, Brook rebuts the claim that these modes can be graded according to status. Even if they could, it is not necessarily the case that e-English must be superior or inferior to speech or writing. One reason is that it is in many cases more rule-free and much less formal than both modes.

In view of the foregoing, one may be tempted to look at e-English as a third, hybrid mode that has affinities and differences with both writing and speech. However one has to wait and see whether this “third” mode will harmonize itself with either of the two older modes or will choose to go further afield and establish itself as an independent mode.

CONCLUSION

This study was meant to investigate the main linguistic features of the English used on the IRC channels and the Y! chat rooms. It focused on sentence structure, i.e. sentence length and sentence complexity, in addition to lexical analysis, i.e., word truncation, orthography, word formation and taboo words. It also examined the similarities and differences between Cyber-English on the one hand and each of written and spoken English on the other.

Analysis of the data indicated that the overwhelming majority of the chatters' sentences were short and simple sentences. Analysis also showed that the e-chatters used several word truncation methods, e.g., using one letter for a whole word and eliminating one or more vowels in a word as long as it is still understandable. They also tended to misspell words either intentionally or accidentally. Further, they applied regular methods of word formation, e.g., derivation, coinage, acronymization, etc. to produce their Cyber-words. Finally, they showed almost no restriction on using taboo words; on average, the corpus included one taboo word for each ten sentences.

The study provided evidence that Cyber-English shares attributes with both spoken and written English. However, it cannot be facilely categorized as either written or spoken

English. It was argued that e-chat English would be better viewed as a newly emerging, hybrid form with its own characteristics and uses.

Since each chat channel is a Cyber-community that can be distinct from other channels. One task that awaits further research is to examine whether the name of a channel/room, e.g., Professors' Chat Room, Islam Chat, etc., has an impact on the linguistic features of the online English used in them. Moreover, this study has been confined to features of e-English. Hence, it could be interesting to examine the linguistic features of chats conducted in other languages. A study of chats in e-French or e-Arabic, for instance, might reveal features that were not detected in this work. Alternatively, such a study might show that the linguistic features of online chat tend to be universal. Finally, at the syntactic level the study focused on sentence length and sentence complexity. Further research may investigate the frequency of occurrence of sentence forms, i.e., declarative, imperative, interrogative and exclamatory, in e-chats and the communicative functions each form conveys. Such a study may disclose useful information on the similarities and differences between online chat English on one hand and spoken and written English on the other.

NOTES

1. This research is based on re-analysis of data collected for the first author's thesis (Al-Sa'di, 2003) conducted at the University of Jordan under the supervision of Jihad Hamdan.
2. Bisexual samples are those that have taken place between two people of opposite sexes. Unisexual samples are those performed by two interlocutors of the same sex. These two terms, i.e., *bisexual* and *unisexual*, as used here, are not at all intended to allude to sexual preferences of the Cyber-chatters.
3. The term *page* as used here has the following features: typeface is Times New Roman or any other font style that does not occupy a different area on a sheet of paper from that occupied by Times New Roman when the font size has not been changed, font size is 14, line spacing is single, and size of the sheet is A4. These measurements in particular have been taken up as they are universally the commonest. However, these norms may be deviated from by e-chatters in terms of font type and size, which often happens and is quite natural, and these chats were duplicated as such to make sure the authentic texts used in this work were recorded in good faith.
4. For a simple test, the researchers asked five native speakers of American English (none of them with any special training in linguistics) to categorize this sentence as either "short" or "long." The invariable answer was: it's a short sentence. On the other hand, Microsoft Word, the best-developed word processing computer program, recognizes by default as "long sentences" only those which exceed 60 words in length.
5. Examples that do not start with an ID or nickname have been taken from a private chat session in the corpus of this study (participants in private chats have been anonymized). If the example starts with an ID or nick, it should be understood to be taken from a public chat session. Moreover, examples from Y! use a colon after the ID; examples from IRC use angle brackets to enclose the nickname. The two posts reported here would be "de-misspelt" thus:
 <A> Why are you on invisible?
 Because I don't want to talk to my ex (-boyfriend).
6. Those quotations of dictionary entries are not facsimiles of the original entries. Irrelevant material in the entry (e.g., phonetic transcription) was omitted. Moreover, when a *taboo* word has more than one sense or entry, only those that directly address our concern here were copied.
7. Even in public chat channels, an e-chatter cannot be viewed as writing to the public but only to a certain number of e-chatters.

REFERENCES

- Al-Sa'di, Rami (2003) An investigation of the linguistic and textual features of "Internet Relay Chat (IRC) and Yahoo Messenger (Y!)" English. Unpublished MA Thesis, University of Jordan.
- Brook, George Leslie (1974) *Varieties of English*. London: Macmillan.
- Collins English Dictionary* (2000) Edited by Jeremy Butterfield. London: Collins.
- Crystal, David (2001) *Language and the Internet*. Cambridge: CUP.

- Dean, Leonard F. and Wilson, Kenneth G. Wilson (1964) *Essays on Language and Usage*. Oxford: Oxford University Press.
- December, J. (1966) CMC Magazine. <http://www.december.com/cmc/mag/1996/dec/shalin.html> (August 16, 2003).
- Freeborn, Dennis (1994) *From Old English to Standard English*. London: Macmillan.
- Halle, Morris (n.d.) The Sounds of Speech. <http://www.lsadc.org/Halle.html> (August 16, 2003).
- Heffernan, James A. W. and Lincoln, John E. (1990) *Writing: A College Handbook*. New York: W. W. Norton & Company.
- Kaplan, Jeff (1989) *English Grammar: Facts and Principles*. London: Prentice-Hall.
- Longman Dictionary of Contemporary English* (1995) Third edition. Harlow: Longman.
- Mizrach, Steve (2003) Sociolinguistics of Electronic Mail. <http://www.fiu.edu/~mizrachs/sociolang-email.html>
[http://www.longman.com/ldoce/Longman Dictionary of Contemporary English](http://www.longman.com/ldoce/Longman%20Dictionary%20of%20Contemporary%20English) (Online) (August 16, 2003).
- Oxford Advanced Learner's Dictionary* (2000) Edited by Sally Wehmeier (sixth edition). Oxford: Oxford University Press.
- Oxford Advanced Learner's Dictionary of Current English* (1974) Edited by Albert Sidney Hornby. Oxford: Oxford University Press.
- Oxford Advanced Learner's Dictionary of Current English* (1995) Edited by Jonathan Crowther. Oxford: Oxford University Press.
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey, and Svartvik, Jan (1972) *A Grammar of Contemporary English*. London: Longman.
- Rinvoluceri, M. (n.d.) Spoken and Written Language. http://www.etprofessional.com/articles/2_104.htm (August 16, 2003).
- The American Heritage Dictionary of the English Language* (2000) Edited by Joseph P. Pickett et al. (fourth edition). Boston: Houghton Mifflin.
- Webster's Dictionary and Thesaurus* (1993) Edited by David B. Guralnik. New York: Simon & Schuster.
- Werry, C. (2003) Computer-Mediated Communication. http://www.msn.edu/lingua_praxis/v01/23587.html (August 16, 2003).

(Received 7 January 2005.)

APPENDIX I: A PRIVATE CHAT SAMPLE

- <A> sup?
- o yeah?
- <A> ok y r my pms on?
- umm, i figured it would b more discreet to perpetuate chat drama in a pm
- <A> ok
- <A> understandable
- it was Melio
- <A> ohe
- and he was a bitch about it
- <A> yea he's a dumbass
- <A> he's always a bitch
- i hung out with him a couple weeks ago, and all he talked about was how he was a chat god
- <A> he went gossippin like a bitch would do
- <A> u there
- <A> r u dc snieg/soz babes lmfao ne case try log in off agen
- it was hilarious
- <A> lmfao@Chat god
- :D
- “i'm going to kick his ass”
- <A> how many times a day u hear that on chat? power of anonymity
- yeah, that's about it 2
- it's funny. i think christian has got the whole game under hi belt
- *his

APPENDIX II: A PUBLIC CHAT SAMPLE

savant: that's true, but does one have to memorise the dictionary by rote?

metonyme: **i see that i have erroneously presumed that that prior statement would need no further clarification**

perrenial_student: its only a problem because communication suffors for it

perrenial_student: it affects everything form social interaction to politics

rnm_70 left the room

liuzhen515: I am from China

metonyme: **well it certainly does not help having a dunce president.**

liuzhen515: who want to talk with me?

anthea0170: i'm from china, too.

rahulsalonia joined the room

shawtboi: I'm an American at Harvard, and I major in Biology and Chemistry, and with the terms we use i would not say America has a dense vocabulary but only common words are understood in society

perrenial_student: I sit in these classrooms whrre a professor asks the students to elaborate on an idea, and they completely understand the concept, but cant vocalize it

cynthiaaah joined the room

metonyme: **doesn't reflect well on the nation**

roses4u joined the room

mehdi_111111 left the room

cynthiaaah: hey roses, roomies

cymbal_c: i'm from havard also

roses4u: morning all

savant: shawtboi, what the layman knows and what standard english is r 2 different things

roses4u: hi cyn

ham5104 is away (Auto-Away)

shawtboi: That's the effect of poor highschool training

shawtboi: I know

shawtboi: Just like ebonics and english

abstract_light left the room

liuzhen515: hi,who want to talk with me?

cymbal_c: i agree

hatami72 left the room

perrenial_student: I had long ago met my humanities requirements but this tech university im at force d me to take their general version of humanities acredidation

metonyme: **i think english remains unnecessarily irregular**