Linguistic Features in the Online Discussion Forums

Z. A. Kadir, M. Maros, and B. A. Hamid

Abstract—**Due to the revolution of technology in communication, a unique form of communication has emerged involving the use of new word forms, structures and style of expression. Therefore, this study examined the linguistic features used in a discussion forum among the e-distance learning students in Universiti Teknologi MARA Malaysia. The data of this study were collected from messages posted by the students to their friends and lecturers. The online forums are platform for the students in this program to communicate with their fellow classmates, lecturers and other e-distance learning students. The study indicates students use linguistic features such as rebus writing, internet specific acronyms, reduplication of letters when communicating with their friends. This study also hopes to enrich the knowledge on the study of computer-mediated communication and to motivate other researchers to conduct further studies in this area.**

Index Terms—**About four key words or phrases in alphabetical order, separated by commas.**

I. Introduction

Computer-mediated discourse is the communication produced when human beings interact with one another by transmitting messages via networked computers. The study of computer-mediated discourse is a specialization within the broader interdisciplinary study of computer-mediated communication, distinguished by its focus on language and language use in computer networked environments, and by its use of methods of discourse analysis to address that focus [1]. It encompasses all kinds of interpersonal communication carried out on the Internet, e.g. by email, instant messaging, web discussion boards, and chat channels.

The Internet allows people to engage in a multi-party conversation online, either synchronously, in real time or asynchronously, in postponed time. Such interactions take place in various forms of Internet communication such as chatgroups, newsgroups, usergroups, chatroom [2],[3]. In asynchronous setting, users are not required to log on at the same time in order to receive and send messages as their messages are sent to a central computer address and inserted into a permanent refreshing screen along with contributions from other users. The online members of the group see their contributions appear on screen after they make them and hope for a prompt response [1], [2],[3]**.** The group members communicate for various reasons such as interest in a particular subject matter, entertainment or communication. The groups may comprise of people from various fields; academic, professional, governmental, commercial or social [2],[3].

In a synchronous setting, electronic interactions take place in real time. Some systems are designed to facilitate communication between at least two users or several users, e.g. IRC. The scenario of two people talking in parallel and at a length while gaining full mutual understanding is inconceivable. With multiple interactions, the situation may be potentially confusing as it is difficult to know how many people are involved, who they are and what they are talking about. This type of setting is mainly associated with nicknames-system. Clashes of nicknames never occur. Topics normally obsolete very quickly and it is not easy to keep track of conversation after some time [2],[3].

Online discussion forum is also a type of asynchronous setting of communication whereby the sender can leave and expect responses to his/her messages. In other words, the users are not only able to discuss and share information, but also exchange ideas and opinions. The discussion forum allows communication with many participants' structure. The discussion forums can be divided into three different components which actually depend on the display format; non-threaded, semi-threaded and fully-threaded. The non-threaded forum function is similar to notice bulletin board as it does not encourage discussion or reply to messages. The messages are normally display in chronological order. The semi-threaded online discussion forums, on the other hand display the initial messages on top and allows other users to reply to the messages. The replied messages will then line up under one another below the initial message which is normally in chronological with the most recent messages display below or vice-versa. Finally, the fully threaded online discussion forums enable other users to reply to the initial messages. The basic layout is similar between the semi-threaded and fully-threaded online discussion forum. However, users in fully-threaded discussion forums are allowed to response to the replies of the initial messages which later on can develop into a string of discourse. This format draws the closest to face-to face discussion that supports turn-adjacency.

II. Background to the Study

E-learning, a type of learning that is considered important in the emerging knowledge of society. Hence, computer-mediated communication has been adopted in most e- learning setting. Computer-mediated communication refers to any form of communication enabled through the use of computers. In the computer-mediated communication literature, the term generally refers to communication modalities dependent on computer-based networks or meta-networks, particularly the Internet and commercial online services. Language plays an important role in the communication activities in the computer-mediated communication. Due to the revolution of technology in

communication, a unique form of communication has emerged involving the use of new word forms, structures and style of expression.

The online discussion forum, an asynchronous mode of communication, is a platform for Universiti Teknologi MARA students who are in the e-distance learning program to discuss academic and non academic matters with their learning facilitators and classmates. The forum has no moderators, but the students are expected to abide to the ethical code published online (politeness and respect for others). Each course has its own discussion forum designed for them. All students who are studying part time, full time or e- distance learning at the Universiti Teknologi MARA are taught in English. Therefore, these students who are taking English courses are expected to use English language either in offline or online communication. However, an interesting phenomenon seems to appear in the online discussion forum among the e-distance learning students; that is code-switching and the use of paralinguistic features. It appears that these students use various linguistic features when they write their messages online in spite of the fact that the discussion forums are designed for English courses and should be written in English. Therefore, the main objective of this study is to examine the linguistic features used in the discussion forum.

III. LITERATURE ON LINGUISTIC FEATURES IN COMPUTER-MEDIATED COMMUNICATION

In computer-mediated communication, writers have to use other manipulation of written signs in order to accomplish pragmatic work that could be achieved through phonological variation, prosody, gesture and other cues in ordinary spoken conversation. Segerstad [4] examined the linguistic feature in computer-mediated communication found that writers use all capital letters, repetition of words, emoticons, asterisk, symbol replacing words to as paralinguistic cues in the interaction. In addition, Crystal [5] in his book, Internet Linguistics writes that text abbreviation is actually not a modern phenomenon. Many of these abbreviations are found in chatroom interactions even before the existence of mobile phone and some of them can be dated a hundred years or more. Moreover, the omission of letter as in msg (message) and xlnt (excellent) is not a new phenomenon. According to him, Wric Partridge published his dictionary of abbreviation in 1942 which contains a lot SMS looking examples such as agn 'again', mth 'month' and gd 'good'. Internet interactions lack the facial expression, gestures and conventions of body posture that are considered important when expressing ideas and opinions. Therefore writers use various ways to express themselves such as the use of emoticon, bold or block letters. However, despite the creativity of the art, the semantic role of emoticon is rather limited. For example, the basic smile can mean sympathy, delight, amusement and others.

Another prominent linguist in computer-mediated communication, Crispin Thurlow [6] studied mobile messages among the first year Language and Communication at Cardiff University Students. Participants were asked to retrieve from their phones 5 messages that they had either sent or received. A total of 544 separate messages were recorded and transcribed. The length of the individual messages was calculated using the standard Microsoft word count function. Based from the investigations, Thurlow asserts that each individual does not have one style of language in any environment, instead, she/he will have a repertoire or a range of style to suit different context. The following common patterns were found;

- shortenings (missing end letters), e.g. 'lang' for 'language'.
- Contractions (missing middle letters), e.g. 'gd' for 'good'
- ;g' clipping (final letter missing, e.g. 'goin' for 'going'
- Other clippings, e.g. 'hav' for 'have'
- Acronyms and intialisms, e.g. 'v' for 'very'
- Letter/number homophones, e.g. '1' for 'one'
- Non-conventional spelling, e.g.'sum' for 'some'
- Accent stylization (speaker tries to represent a particular pronunciation, for example regional speech), e.g. 'wivout' for 'without'
- Non-alphabetic symbol
- Emoticon

(page 19)

Based on the above findings, Thurlow concludes that a number of sociolinguistic maxim or triggering factors are required to explain some of the features above:

- Speed-txters have to speed up their pace of communication, so they need to take short cuts.
- Brevity-txters have only limited space for their communication, so they need to omit any elements that are not strictly necessary for understanding
- Paralinguistic restitution-txters need to find ways to replace the aspect of physical communications such as body language that are absent
- Phonological approximation-txters want to build in ways their readers to 'hear' their voice, so try to change the written language to represent this.

Although research conducted by Thurlow was based on mobile text messages, there were some similarities with the paper in her research and the pattern of messages in this study. In this research, participants also used emoticon, non-conventional spellings and clipping when writing their messages online. Therefore, Thurlow's text patterns and Danet's [7] general characterization of digital discourse provide a useful frame of reference. Thus, as mentioned by Crystal [5], these styles of writing exist before the invention of mobile phones.

Another study on SMS messages was conducted by Anis [8]. A total of 750 French messages were collected from four volunteers. Based from his research, he categorized the corpus into three broad types: phonetic spelling, syllabograms (rebus writing) and logograms (symbols, unilateral abbreviations, acronyms) (page, 97). In the phonetic spelling, he discovered the texters not only substituted, reduced vowels or consonants but also deleted silent letters in their messages. For example, the substitution of 'z' for 's' ("pleaze' for "please"). Another striking feature was syllabogram or rebus writing such as the use of a letter of or a number to represent the phonetic sequence that constitutes its realization in spoken language such as 'b4' stands for 'before'. The third finding in his research was logograms which involved not only word signs such as "@"

for 'at' but also single-letter abbreviations such as "CNN" (Cable News Network) (page, 105). He concludes that such messages are intentional, creative and definitely comprehensible to their recipients. In addition, the messages also reflect common human characteristics.

Norizah Hassan and Azirah Hashim [9] in their studies of the features and language use in electronic English in Malaysia highlight how language has been used creatively online by different ethnic groups in Malaysia. The data was taken from a corpus of 2 million words collected from various electronic genres: blogs which are written for informal readers, chats from Malaysian chatrooms, instant messages, emails and text messages between friends. In their preliminary finding, many features of spoken Malaysian English as well as other varieties of linguistic features are found in the online communication. Intersentential and intrasentential code-switching occur between English and Malay, Chinese dialects, Tamil and Iban. According to the above researchers, the features are commonly found in spoken Malaysian English except the use of symbols like @, the use of emoticons for expressions and use of the Roman script to represent sounds in Chinese. Internet users also establish their identity through the use of features specific to the variety and through the medium that is used. The study offers a general overview of the use English on the Internet.

IV. METHODOLOGY

The primary source of the study is discussion forums taken from the e-distance learning program, Universiti Teknologi MARA Malaysia. Students who are taking diploma programs will have to go through a minimum of three semesters of English courses depending on the programs they are taking. All the first semester one students are therefore to take a proficiency course, BEL120 Consolidating Language Skills.

A. Participants-Students

Participants for this research were the students of Universiti Teknologi MARA studying in the e-distance learning program in Shah Alam campus. The participants were the first semester students in Diploma in Business Management. The postings indicate that participants code-switched between Malay and English in the same posting or post complete messages in either Malay or English only. Some participants were more active on the threads as compared to others.

B. Posting collection

A total of 110 messages were collected from the Diploma of Business Management students from July 16 2009 till Oct 27 2009. Each course in the e-distance learning program is provided with its own discussion forum platform, thus the data was then collected from BEL120, an English course for the first semester students. The content of the messages were mainly about classes, exams, documentation, assignments to be submitted and rarely about everyday life. The e-learning students only attended seminar four or five times per semester, therefore, they had to communicate online with their friends and learning instructors. Although there was no specific amount of posting or messages for them to send in the online forums, it was considered important part in the syllabus of BEL120 as postings sent were given marks for participation. The allocation of 10% of the total marks was given for attendance and online participation.

The posting messages collected were then divided into 4 categories of postings; Malay posting, English posting, code-switching and others. Malay postings consisted of postings that were written in Malay language only, while English postings consisted of English language only. Code-switching posting consisted of intersentential code-switching and intrasentential code-switching between Malay and English, English and Arabic, Arabic and Malay. The category 'others' on the hand consisted of messages posted in other languages or symbols only. The largest number of postings was from and the English language category, followed by the category of code-switching between Malay and English language posting. The least number was from 'others'. It is not surprising that English language postings contribute the largest number of postings as the postings were taken from the BEl120, an English language course forum.

TABLE I: TOTAL NUMBER OF POSTING MESSAGES AND DATE

GROUP	No. Of Posting	Date
EBMS1B	110	1 July 2009-27 October 2009

The messages were collected for more than 3 months and the total number of posting collected was 110 messages. Students sent 86 posting messages while learning facilitator sent 24.

TABLE II: TOTAL NUMBER OF POSTING SENT BY STUDENTS AND FACILITATOR

Group	Students	Learning facilitator
EBMS1B	86	24

TABLE III: TOPICS DISCUSSED IN THE ONLINE DISCUSSION FORUMS

Group	No of topics	Topics
EBMS1B	15	BEL120 Final Exam Exam Grammar Test 4th Seminar Essay assignment Answers to Past Semester Questions 3rd Seminar Assignment 2nd August Answers to Part B Portfolio Longman Dictiionary 2nd Seminar Answers to 1st Seminar quiz 1st Seminar

V. DISCUSSION

A. Letters/Numbers Homophones

Extract 1

M: still can't believe that i will attend an examination **2moro** since left high school almost 12 years-sooooo nervous., haiyoooo...

M:wish me luck ehhh...n **2** all my frennnzz good luck 4 u'll...hope that we will **2**gether next sem..insya-al1ah

In extract 1, M wrote that she would be taking an examination after leaving school for quite some time. Instead

of writing the complete word 'tomorrow', the student wrote '2moro'. The word 'to' and '2' is what is considered homophone, the same sound in English. In the excerpt above, M seems to replace the word 'together' and 'to' with '2'. In addition, M also wrote '4' instead of preposition 'for'. Anis [8] describes the use of a letter or a number to represent phonetic sequence that constitutes in spoken language as rebus writing or sylllabogram or rebus writing. He also adds that the use of syllabograms allow a number a drastic reduction of the number of characters.

B. Eccentric spelling/Accent Stylization

Extract 2

S: Salam puan...sooo nervous and still shaking., thanks for your advice..

Extract 2 shows how the writer expressed her feelings of nervousness as the final exam was approaching. In eccentric spelling or accent stylization, letters are reduplicated. The example given shows the reduplication of vowel 'o'. In other case of such feature, words are spelt in what is called 'pronunciations' spelling [10] such as when writer used the word 'coz' for 'cause'.

C. Capital Letters

Extract 3

L: Grammar Test will be conducted this Sunday (11/10/2009). EVERYONE MUST TAKE THE TEST ON THAT DAY OR YOU WILL LOSE 25% OF YOUR OVERALL MARKS. THERE IS NO MAKE UP TEST. SO SEE YOU THIS SUNDAY.
   HAPPY STUDYING/ V'

The writer of this extract is actually the learning facilitator of the course. It is interesting to note that the use of capital letters in the discussion forums were mainly from the learning facilitator or lecturer. According to Nishimura [10], users choose to use capitals letters most probably because they are visually more prominent as compared to lower case. In the example given, the learning facilitator reminded the students that they must take the exam on the specific date or they would risk losing 25% of the overall marks for the course. Therefore, the use of capital and bold letters function as warning and reminder for the students the importance of the test.

D. Written out Laughter

Extract 4

J: Salam Pn r. I also would like to order the dictionary too. But i forgot the price that u have been mentioned it. I pay you twice boleh?? Hehehe

As mentioned, in computer-mediated communication, writers have to use other manipulation of written signs in order to accomplish pragmatic work that could be achieved through phonological variation, prosody, gesture and other cues in ordinary spoken conversation. Therefore, the feature of written out laughter help to replace the missing of other vocalizations and nonlinguistic sound/noise. Extract 4 displays how the writer was making jokes with his learning facilitator regarding the method of payment for purchasing dictionary. In the messages, P was asking if he could pay the dictionary in installment.

E. Emotion/Emoticon

Emoticon characters may consist of colon, parentheses and other characters. Writers can create face masks or copy and paste from ready-made marks. In discussion forums, writers can change text emoticons to images that are called 'emoticons'. Emoticons are normally used to express writers' feelings or mood. In this study, writers use smiley faces to show their emotions. Even though emoticons are popular, writers do not always include them in their messages. Nishimura [10] asserts that users tend to employ more emoticons in synchronous interaction as compared to asynchronous interactions.

F. Clippings

Extract 5

N;ok...noted..!! wat 'bout my 1st assgmnt?

Extract 6

J: tanks

In the above examples, writers deleted the silent 'h' the word 'what' and 'tanks'. The words that the writers used is intelligible and comprehensible although it violates the rule of standard English spelling Clippings also help writers to write their messages faster.

G. Other Linguistic Features

Code-switching

Gumperz [11] defines code-switching as the alternation in the form of two consecutive sentences in which a speaker uses the second language as a reiteration or reply to someone's statement. He considers code-switching as a type of contextualization cue, defined as 'speakers' and listeners' use of verbal and nonverbal signs to relate what is said at any one time and in any one place to knowledge acquired through past experience, and in order to retrieve the presuppositions, they must rely on how to maintain conversational involvement and assess what is intended'' [11]. Contextualization cues do not carry meaning in and of themselves. They convey meaning only when they co-occur with verbal cues in particular moments of interaction.

A. Code-switching for Formulaic Discourse Purposes:

- Greetings

Extract 7
   A: Salam puan (Peace be with you madam)
   B: Assalamualaikum Puan R (Peace be with you)
Extract 8
- Farewells
   C: Wassalam (Regards)
   D: Insyaallah (God willing)

The majority of students in this university are Malay and of course being a Muslim, some Arabic expressions are inserted in the beginning of the posting. The word 'salam' is a short form of 'assalamualaikum'. The meaning of assalamualaikum is equal to 'peace be upon you'. It is common in Muslim society to begin their conversation by giving 'salam' when they meet each other or even when they start their conversation over the phone. It is interesting to note that students in the discussion forum use Arabic words especially in the salutation and closing of their messages. The switching in Arabic may also indicate that the students are displaying their Muslim identity in their communication.

B. Switching to Show Respect

Extract 10

M: puan (madam)

Extract 11

T: Noted pn (noted madam)

In this study, the use of expression 'puan' which means madam can be found in 53 messages out of 86 messages posted by the students. In Malay culture, students are expected to respect their teachers as they are the people who teach and give you the knowledge regardless of their age. It is interesting to note that the majority of the messages posted clearly shown the students respect by specifically addressing the learning facilitator as 'puan' which reflects the culture of the Malay society.

C. Switching to Mark a Specific Terminology

Switching is also used to mark a specific terminology especially when it is connected to the use of computer-application. Jayatilal [12] says that the insertion of English lexical items in Malay sentences could summon up the image of the write. In addition a bilingual tends to mix elements from the languages commonly associated with that topic when she/he wants to sound knowledgeable or precise in discussing a specific topic.

Extract 11

S: *Dah ada la F...tak dapat bkk ke?...saya dapat. Dan* save *pun.*

S: *It's already there F, didn't you get it? I got it and saved it. j*

Y: *cuba kak tengok* link *ni:j*

Y:try to look at this link

In the example above, both students started their posting in Malay. However, they inserted English terminology' link and save' instead of Malay word 'simpan' and 'pautan' when the terms used are connected to computer terminology. Other terminology that inserted in their writing are download, email, upload.

D. Switching with Repetition to Emphasize or Clarify Message

Extract 12

Y:*yang 10 soalan ni ke*??? just exercise *kan? yang ni ke*?

Y this 10 question right? Just exercise right? is this the one?

This was also a reply posting to the same posting as the switching for emphatic purposes. Y replied to the posting by asking whether the exercises given by the learning facilitator was the correct one. Y repeated the same question three times started with English, then switched to Malay and finally ended with Malay. This is to emphasize and clarify that he/she managed to get the questions correctly.

VI. CONCLUSION

The study describes the creative ways students use when communicating online with their friends. However, the nonconventional spelling used actually raises educational concerns as teachers are worried that such writing might appear in their formal written document. However, further studies are clearly needed in order to justify the claim.

The samples in this study also show that code-switching is used in similar ways to those of face-to-face conversation.

Primarily, the use of Arabic lexical item in the discussion forum indicates two important things; culture and religion. Firstly, it reflects their identity as Malay and being a Malay also means that the person is a Muslim. In Malaysia, once a baby is registered in the birth certificate as Malay, then she/he is also automatically registered as Muslim. It is something that is common among Malay to greet each other in Arabic although they might not know Arabic. In fact, Malay children are taught to greet others in such a way since they are small. Malay and English Code-switching serves other discourse functions. It is used to mark a formulaic expression and specific terminology. It also used a strategy to capture readers' attention and helps emphasize or clarify a message. Therefore, it can be said that that code-switching in this study occurs intentionally and not due to lack of competence in the base language.

In conclusion, although this research is primarily on online discussion forum, the evidence shows that participants use various linguistic features when they write their messages online. I would like therefore to recommend that future research on code-switching in Malaysia be carried out on the comparison and contrast between the synchronous and asynchronous mode of communication.

REFERENCES

[1] S. Herring, "Computer-mediated discourse," in *Handbook of Discourse Analysis*, D. Tannen, D. Schiffrin, and H. Hamilton, Eds. Oxford: Blackwell, 2001, pp 612-634.

[2] D. Crystal, *Language and the Internet*, Cambridge: Cambridge University Press, 2001, ch.1.

[3] D. Crystal, *Language and the Internet*, 2nd edition. Cambridge: Cambridge University Press, 2006, ch.1.

[4] Y.H. Segerstad, *Use and Adaptataionof Written Language to the Conditions of Computer-Mediated Communication, Ph.D.* dissertation, Goterborg University, Sweden, 2002. Available: http://www.ling.gu.se/~ylvah/dokument/ylva_diss.pdf/

[5] D. Crystal, *Internet Linguistics*, London: Routledge, 2011, ch.1.

[6] C. Thurlow, "Generation Txt? The sociolinguistics of young people's text-messaging," *Discourse Analysis Online, 1* (1), pp1-27, 2003. Available: http://faculty.washington.edu/thurlow/papers/Thurlow(2003)-DAOL

[7] B. Danet,(2001), Cyberpl@y: Communicating Online, New York: Berg companion website. http://pluto.msc.huji.ac.il/-msdanet/cyberpl@y

[8] J. Anis., "Neography: Unconventional spelling in French SMS text messages," in *The Multilingual Internet*, B.Danet and S. Herring, Eds. Oxford: Oxford University Press, 2007ch. 4.

[9] N. Hassan and A. Hashim, "Electronic English in Malaysia: Features and language in use," *English Today*, vol 25, No 4, pp39-46, Dec.2009.

[10] Y. Nishimura, "Linguistic innovations and interactional features in Japanese BBS communication," in *The Multilingual Internet*, B. Danet, and S. Herring, Eds. Oxford: Oxford University Press, 2007,ch. 7.

[11] J.Gumperz, *Discourse Strategies: Conversational Code-Switching.* Cambridge: Cambridge University Press, 1982, ch.4, pp59-99.

[12] R. Jayatilal, "Code-switching and transfer in email correspondence exchanged between four Malaysian women," *Jurnal Pendidikan Tiga*, 1998, Jilid 2; Bil 2. Available: http://www.scribd.com/doc/15339217/Codeswitching-in-ESL

Zaemah Abd Kadir is a senior lecturer at the Academy of Language Studies, UiTM Shah Alam. She holds a Masters Degree in Applied Linguistics and TESOL from University of Leicester, UK. Currently, She is pursuing her doctoral studies at the Universiti Kebangsaan Malaysia.

Marlyna Maros is an Associate Professor at the School of Language Studies and Linguistics, Faculty of Social Sciences and Humanities, UKM. Her research areas include contrastive linguistics, contrastive pragmatics, discourse analysis and sociolinguistics. Her Ph.D thesis is on the speech act of complaint in Malay.

Associate Professor Dr. **Bahiyah Dato' Hj. Abdul Hamid** is a lecturer of the School of Language and Linguistics, Faculty of Social Sciences, Universiti Kebangsaan Malaysia. Her areas of specialization are in Sociolinguistics, Language and Gender and Applied Linguistics. Dr. Bahiyah's research interest is in linguistics mainly sociolinguistics and applied linguistics with special focus on code alternation and code choice, language and gender, discourse analysis, language awareness and identity construction