

# **Is the Chatting between Students Spoken Form, Written Form or Something Else? -Analysis of KWCCDLP (Korea University Waseda University Cross-Cultural Distance Learning Project) Data-**

**Nari Lee  
Korea University**

This paper deals with chatting data of Korea University Waseda University Cross-Cultural Distance Learning Project (KWCCDLP). The main concerns are the analysis of synchronous chatting data and showing its style. It can be spoken or written form or something new. It starts to show the definitions and categories of spoken and written styles from the previous studies. Then these will be compared to the analysis of KWCCDLP chatting data.

The results show that the chatting data has spoken, written and chatting forms which are well mixed. The ones prefer to use spoken forms also have tendency to use more chatting forms than those of preferring to written forms.

## **1. Purpose**

As the computers and internet access become common in schools and homes, the concern of Computer-Mediated Communication (CMC) is increased. It has also brought many changes into education, especially English learning. It helps the students who learn English as a foreign language to have more chances to use English in their communication.

CMC is consisted of synchronous CMC (e.g. chatting or Net meeting) and asynchronous CMC (e.g. e-mail). The methods and effect of CMC have been widely studied. Many researchers (Park, Nakano & Lee, 2003; Harris & Wambeam, 1996; Hubbard, 1997; Takayoshi, 1996) have already compared CMC based education with traditional one. Yet in synchronous CMC, the analysis

of chatting style has hardly been studied.

The main purpose of this paper is synchronous CMC discourse analysis from KWCCDLP data. And it will be categorized into three sections: spoken, written and chatting.

## **2. Literature Review**

According to Park, et al. (2001), discourse is the result of communication including conversation, interview, written text and so on. Discourse analysis is the study of spoken or written communication.

Discourse is divided into spoken and written discourse (Kaplan & Grabe, 2002; McCarthy, 1991; Stenström, 1994). The pauses, in other words, tone unit boundaries mark the spoken discourse while punctuation marks do the written discourse (Stenström, 1994). Chafe & Danielewicz (1987) use lexical usage. They claim that spoken discourse has less and limited lexicons than written one. Similarly, Corson (1997) states the lexicons used in written discourse are more various and newer than those in spoken discourse. In aspect of lexical density, written discourse has a higher lexical density than that of spoken (Ure, 1971). Halliday (1992) also has similar study of lexical density. He argues that lexical density of written discourse has twice more than that of spoken. And written discourse has more lexical items, while spoken discourse has more functional items.

Spoken discourse is here-and-now activity. Speakers take turns and collaborate. To show collaboration, they use backchannels, feedback and hesitation phenomena such as silent, filled pause, verbal fillers, false start, repetition and incomplete utterance (Stenström, 1994). In addition, paralinguistic features like facial expressions and gestures are also used in spoken discourse.

Written discourse has relatively longer and complex sentences and its grammar is more complicated than spoken discourse (Chafe &

Danielewicz, 1987). As Stenström (1994) mentions written discourse had punctuation marks. Using capital letters is another distinctive feature of written discourse.

On the other hand, some claim that it is hard to distinguish these two (Kaplan & Grabe, 2002; Biber, 1988). Many of features in spoken and written discourse are overlapped. For example, writing a letter to a friend is similar to spoken discourse like face-to-face conversation despite using pen and paper with writing skills. Having a lecture in a class, however, is more like written discourse despite using verbal expressions. It uses more complicated words and professional terms than every day conversation.

CMC discourse also has features of both of spoken and written discourse. In synchronous CMC discourse, participants type messages using keyboard like written discourse and they interact simultaneously like face-to-face spoken discourse. Graddol (1989) and Johanyak (1997) claim that synchronous CMC is hybrid forms of communication having both features of spoken and written discourse. On the contrary, Collat & Belmore (1996) mention CMC discourse:

“Messages delivered electrically are neither ‘spoken’ nor ‘written’ in the conversational sense of these words. There is an easy interaction of participants and alternation of topics typical of some varieties of spoken English. However, they can not be strictly labeled as spoken messages since the participants neither see nor hear each other. Nor can they be considered strictly written since many of them are composed directly on-line, thereby ruling out the use of planning and editing strategies which are at the disposal of even the informal writer.”

But Collat & Belmore (1996) only consider the medium of communication not patterns used in the discourse. For qualitative analysis, Doell (1998),

Werry (1996) and Lim (2003) suggest four subsections of analyzing CMC discourse. These are lexis, orthography, grammar and discourse. Lexis section includes analyses of abbreviation, onomatopoeia, emoticon, slang and jargon. In orthography section, analyses of misspelling, varieties, alternative spelling, capitalization, punctuation, omission and contraction are included. The omission of subject or verb, disagreement of tense, mismatch of third person singular and verb and use of 'ain't' are the components of grammar section. Discourse section includes turn-taking and repetition of short sentences.

### **3. Method**

#### **3.1 Subject**

Among twenty seven Korea University (KU) students who enrolled Global English through Internet I, the chatting data of ten students who had regularly chatted with Waseda University (WU) students were chosen. Two students from KU were consisted in one Korean group and two students from WU were consisted in one Waseda group. And each group was matched for a chatting pair. Simply if ten Korean students' chatting data were collected there were 40 participants. But some students were in the same group and consequently 37 students were participated in. All KU students were native Korean speakers and all WU students were native Japanese speakers. There were 18 Koreans and 19 Japanese students. There were seven males and eleven females in Korean group, eleven males and eight females in Japanese group. The range of age was 18 to 26 years. Basically four students who were pre-matched by TA of each university were supposed to have a chat in a group. However some chats were held with only two students or different students from other groups every week. Among ten chatting data, nine KU students whose chatting data were not overlapped were chosen and numbered from K1 to K9. Following table shows the participants' sex and the number of partners.

Table 1 Participants' sex and the number of partners

participants	sex	number of partners				
		total	KU		WU	
			male	female	male	female
K1	M	3	0	0	1	2
K2	F	1	0	0	0	1
K3	F	2	0	1	1	0
K4	M	3	1	1	0	1
K5	M	2	1	0	1	0
K6	F	1	0	0	0	1
K7	M	10	1	3	4	2
K8	F	1	0	0	1	0
K9	F	6	1	1	3	1
total (M:F)	4:5	29	4	6	11	8

(M: male, F: female)

### 3.2 Procedure

All students who enrolled Global English through Internet I had participated in KWCCDLP. There were video conferencing every week during the class from May to June between KU and WU. Conference topics were decided by students from both universities in the beginning of video conferencing session. After video conferencing, all students were required to have a chat at least once in a week with their partners. Chatting topics were as same as those of video conferencing. Students were required to have a chat for five weeks at multimedia education room where special programs and equipments for chatting were set up. There was more than one TA who helped students participated in chat session. In

the beginning of a semester, there was a trial session for students how to use chatting program, BizMate. With web camera, participants see their partner on the screen. The message are typed with keyboard and word files, presentation files, pictures, graphs and web site can be displayed on the screen while interact. Students were asked to save all chatting data after their chatting. At the end of the semester, students had to submit their chatting data which were downloaded diskette to TA.

## 4. Result and Discussion

### 4.1 General analysis of chatting data

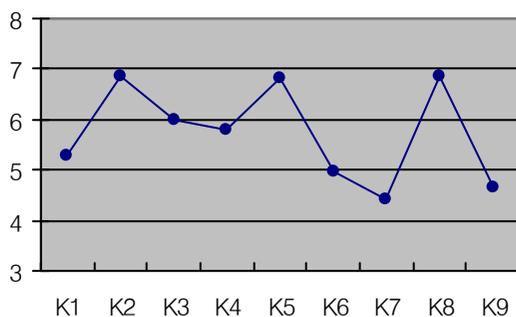
For quantitative approach, the total number of turns and words used in chatting data were calculated. Then the total number of words was divided by the total number of turns. It will be called 'Ave' in this paper to explain which had denser and longer sentences. According to Ure (1971) and Halliday (1992), written discourse has higher lexical density than spoken discourse. Like studies of Ure (1971) and Halliday (1992), if one's chatting data has higher Ave than that of the other, it may be relatively close to written style. Calculating Ave is for comparing quantitative analysis to qualitative analysis. Table 2 and figure 1 show the total number of turns, words and its Ave. The average Ave of all participants was 5.40.

Table 4 Total number of turns, words and Ave of the participants

participants	Turn	Words	Ave
K1	651	3431	5.27
K2	835	5719	6.84
K3	890	5344	6.00
K4	1504	8703	5.78

K5	177	1207	6.81
K6	987	4924	4.98
K7	1882	8289	4.40
K8	587	3996	6.86
K9	1443	6731	4.66
total	8956	48343	5.40
average	218	1179	5.40

Figure 1 Participants' Ave



(average Ave of all participants: 5.40)

Having qualitative analysis, it can be assume that the chatting data which has higher Ave than the average Ave is close to written discourse. On the contrary, the chatting data below the average Ave will show more spoken forms in their data. From table 2 and figure 1, there is an assumption that the data of K2, K5 and K8 have more written forms than the others. K1, K6, K7 and K9 use more spoken forms in their chatting data. For qualitative approach, the spoken, written and chatting forms in chatting data were examined.

#### 4.2 Criteria of chatting data

Doell (1998), Werry (1996) and Lim (2003) mention the subsection

of synchronous CMC discourse. And the features of the subsections are categorized to spoken forms, written forms and chatting forms.

Firstly, spoken forms are consisted of slang, jargon, onomatopoeia, capitalization, overuse of punctuation marks (exclamation, question mark), omission of subject (I, it), omission of verb (be, have), mismatch of tense, number, third person singular verb, turn-taking and repetition of short sentences.

Secondly, written forms are consisted of capital letters (proper nouns, titles, first person singular I, interjection O, the first word of every sentence) and punctuation mark (comma, period).

Lastly, in synchronous CMC discourse, participants need to type their messages as speedy as possible to cope with their interact speed like face-to-face interaction. The chatting forms are consisted of abbreviation, computer jargon, emoticons, symbol and number, varieties and alternative or omission of spelling. There are two types of abbreviation. One is using first letters of each word like 'BTW' which stands for 'by the way'. The other is only using consonants of a word like 'PLS' for 'please'. Participants use computer jargon like 'spam mail' or 'web cam'. The most distinctive feature of synchronous CMC is emoticon. To express participants' emotion and facial expression, they use letters and symbol keys and it is called emoticon. It gives paralinguistic help during the chat session. Participants also use symbol and number to reduce time of typing. For example, 'at' is replaced to '@', some preposition such as 'to' or 'for' are changed to '2' or '4'. Some use number and letters to type a word. For instance, 'tonight' becomes '2nite' and 'skater' becomes 'sk8er'. Alternative spelling is changing some spellings in a word as being pronounced or in a simple way. People use 'boi' as 'boy' and 'dunno' instead of 'don't know'. Varieties are very similar to alternative spellings but they are changed intentionally. Most of varieties are simple forms of longer and complicated words. For example, 'because' is changed to 'cos' or 'cuz', 'you are' to 'u r' and 'people' to 'ppl'. Omission of spelling happens when the

omitted letter cannot affect the pronunciation or meaning. It is often shown at the end of the words like ‘nothin’ or ‘goin’’. Or it happens in the middle of the words like ‘wat’ for ‘what’ and ‘havnt’ for ‘haven’t’. But in this paper, varieties include alternative or omission of spellings. Because they are displayed in very similar ways and it is hard to distinguish participants’ intentions in every changed forms.

Based on these categories, the criteria for spoken, written and chatting forms are made. Since written form has only two features, the analyzing scope of features was settled for two characteristic features in each criterion for logical counting and comparing. Two features are selected which are frequently shown in the chatting data. Table 3 suggests two distinctive features of each criterion.

Table 3 Distinctive features of criteria

Spoken Form	Written Form	Chatting Form
-onomatopoeia (sounds of laugh, interjection, filled pause, repetition of vowels in the middle of a word or consonants at the end of a word, ~mark) -overuse of punctuation marks (.../??/?!!/??!!)	-capital letters (proper nouns, titles, first person singular I, interjection O, the first word of every sentence) -punctuation mark (comma, period)	-emoticons -varieties (including alternative or omission of spellings)

#### 4.3 Spoken form

Two features of spoken form are use of onomatopoeia and overuse of

punctuation marks. They are frequently used to express sounds and intonation. The onomatopoeia includes sound of laugh, interjection, filled pause such as ‘um...’, ‘hmm...’, use of ~ mark and the repetition of vowel in the middle of a word and consonant at the end of a word like ‘hiiii’ for ‘hi’ and ‘verrrry’ for ‘very’. Farfelder (2000) insists the repetition of specific letters in a word be considered as spoken discourse for showing emotions. Using ~mark during the chat is very unique form in KWCCDLP data. It is used for illustrating the intonation of words and sentences. The more use of ~mark, the stronger and the longer intonation is made. Another feature is overuse of punctuation marks. It is also used to demonstrate the emotions of participants. The more punctuation marks are used, the stronger emotions are expressed. One unique pattern in KWCCDLP data is over use of period. It makes filled pause, implies interacts are kept going and illustrates participants’ opinion indirectly. The followings are the examples of spoken forms.

Example 1 Overuse of onomatopoeia and punctuation mark

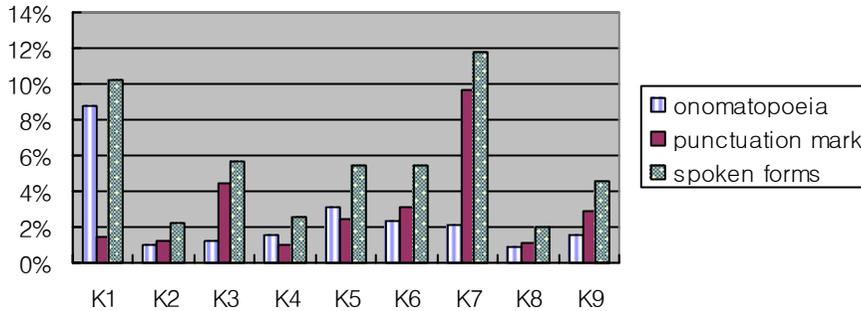
- korea006 : cool~~~~ but appearance is not coolooooooooo~ ha?
- korea001 : inlineskate+\_+wow cool~
- korea001 : nono~
- korea006 : my nick name is ... you can say just "yong"
- korea006 : yong~~
- waseda edu006 : Are you freshman?
- korea006 : wow!!!!
- korea006 : thanks alot!!!!!!!!!!!!!!!!!!!!
- korea001 : kk
- korea006 : that's really reaallly~~ nice complementary to me~
  
- korea006 : yes...that's what we know...
- waseda edu013 : Yes.....Next week is our last chat.....

korea006 : i guess time is up for today...

Two Korean students in the first example use various spoken forms. They repeat specific letters in words like ‘coolooooool’ or ‘reaallly’ to show their strong emotion. Next, they use ~mark to demonstrate intonation, it is mostly used at the end of the sentences or after the words they want to emphasize. There are overuse of punctuation marks on line 7, 8 and the second example. Repeated exclamation marks illustrate amazement and pleasure. Repeated periods show sadness for closing chat session.

Figure 2 demonstrates the frequency of spoken forms used by each participant. It is calculated that the total number of spoken forms including use of onomatopoeia and overuse of punctuation marks is divided by the total number of words.

Figure 2 Frequency of spoken forms used by each participant

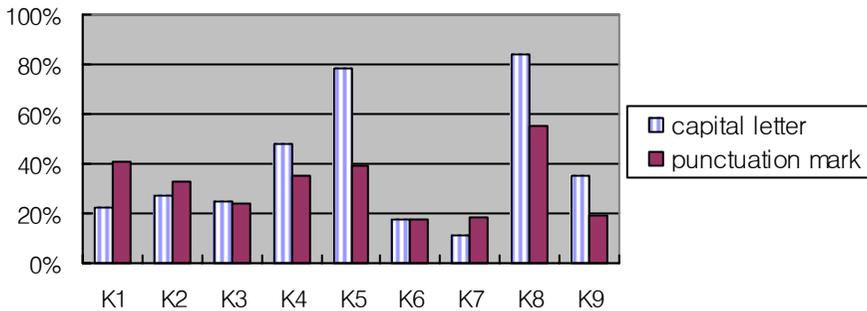


According to figure 2, K7 has the most frequent use of spoken form and K1 has the second most. It is corresponded to the Ave analysis. As the Ave is below the average, it has more spoken forms. In the previous chapter, K1, K6, K7 and K9 are below the average; K1 and K9 have higher use of spoken forms. In contrast, K2, K5 and K8 have higher Ave and their frequencies of spoken form use are relatively very low. Even though K4 and K6 are below the average of Ave, they don't have many spoken forms. It is drawn that there is a tendency that the data with lower Ave are likely to have more spoken forms but not always.

#### 4.4 Written form

Starting each sentence with capital letter and ending with punctuation mark are features of written forms. Figure 3 shows the number of used written forms in chatting data. Being guessed, K5 and K8 which have higher Ave demonstrate more use of written forms. On the contrary, K7 has the least use of written forms. K1, K6, K7 and K9, the ones which have lower Ave usually use less written forms. Although K2 has higher Ave, its use of written forms was low. K2 use less both of spoken and written forms. Except K2, there is a tendency that chatting data with higher Ave has more written forms and lower one has less.

Figure 3 Ratio of using written forms to the total number of turns



#### 4.5 Chatting form

The main characteristic features of chatting forms are emoticons and varieties. Emoticons help participants have paralinguistic cues such as facial expressions or gestures. In KWCCDLP, the participants can see their partners on the screen by web camera. The frequency of using emoticons might be less than the chatting without seeing partners on the screen. Following example demonstrates the use of emoticons.

### Example 2 the use of emoticons

korea008 : ah, did you send?? thanks thans!! ^-----^\*

waseda law006 : yourwelcome

waseda law006 : ^^

waseda edu013 : (@ @)

korea012 : have you ever eaten korean food?

The participants use symbol keys to show their feelings. They use various types of emoticons according to the situation and their feelings.

The next feature is varieties including alternative or omission of spellings. To save time of typing and interact simultaneously, participants use varieties. One way of varieties is typing words as being pronounced like ‘c’ for ‘see’ and ‘foriners’ for ‘foriegners’. Another way is changing spelling of a word like ‘wuz’ for ‘was’ and ‘hab’ for ‘have’. In this case, the pronunciations of words are changed. The last one is omission of punctuation mark apostrophe ’ and changing some spellings like ‘thaz’ for ‘that’s’ and ‘dunno’ for ‘don’t know’.

### Example 3 the use of varieties

korea004 : i hab to go to class and u?

korea004 : c ya next time~

korea004 : hab a nice day~~

korea004 : Thatz okay~

korea004 : because Korean ppl think Nationalism is very important

korea011 : really it wuz.!

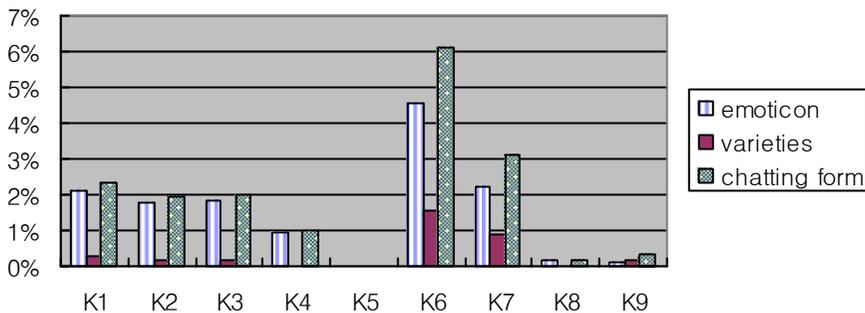
waseda edu012 : always foriners think

korea012 : +\_+ dunno @\_@

korea012 : yes, you r right

In figure 4, there is frequency of using chatting forms. K6 and K7 have higher use of chatting forms than the others. Both of K6 and K7 use many of emoticons and varieties. Generally the use of emoticons is more than that of varieties. The affective filter can be one reason why the participants prefer to use emoticon to varieties. All participants are non native English speakers and they might have tendency to avoid misspelling which shows lower English level. And this happens to the ones who prefer to use written forms in their chatting. For example, K4, K5 and K8 never use varieties and K5 doesn't have any chatting forms. The ones whose Ave is lower than the average of Ave are K1, K6, K7 and K9. The three fourth of these participants have higher frequency of using chatting forms. Among the ones whose Ave is higher, K5 and K8 hardly have chatting forms. But K2 has relatively higher frequency of using chatting forms than those of spoken or written forms in her chatting data.

Figure 4 Frequency of using chatting forms



## 5. Conclusion

According to data analysis, several tendencies of chatting styles are found. Firstly, the chatting data with higher Ave have more written forms than spoken or chatting forms. The other ones with lower Ave have more spoken and

chatting forms. Not always these relations are proven, there is a higher Ave chatting data which has neither spoken nor written forms predominantly but chatting forms.

Secondly, the chatting data which has more spoken forms also have more chatting forms. But the ratio between use of spoken and chatting forms are different. The frequency of using chatting data is relatively lower than that of spoken. It can be explained by KWCCDLP chatting situation and affective filter. The participants of KWCCDLP can see each other on the screen and paralinguistic cues like facial expressions or gestures were delivered on-line situation. With this help, the need of using emoticons is relatively reduced. The other reason is affective filter to use varieties. As mentioned, all participants are non native English speakers and they might hesitate to use varieties. Since varieties change or omit the spellings which are considered as lower English proficiency.

Lastly, the spoken, written and chatting forms are displayed in all chatting data except K5. It can not be concluded that chatting data is spoken style, written style or chatting style by one dimensionally. All features are well mixed and there are only tendencies that one chatting data is close to spoken style, another is close to written style and the other is close to chatting style.

This study is only dealt with the chatting data of 37 non native English speakers for six weeks. For further study, more chatting data are examined and compared to those of native English speakers. The relationship between chatting style and pedagogical implication is needed for the next step.

## Reference

- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Chafe, W. & Danielewicz, J. (1987). Properties of spoken and written

- language. In Horowitz, R. & Samuels, S.J. (Eds), *Comprehending oral and written language*. New York: Academic Press.
- Collot, M., & Belmore, N. (1996). Electronic Language: A new variety of English. In Herring, S. (Ed), *Computer-Mediated Communication: Linguistics. Social and Cross-Cultural Perspectives* (pp.13-28). Amsterdam: John Benjamins.
- Corson, D. (1997). The learning and use of academic English words. *Language Learning*, 47, 671-718.
- Doell, W. (1998, May 11). *The internet relay chat (IRC): Linguistic perspectives*. Wernfrid Doell's personal homepage: Germany. Retrieved June 2<sup>nd</sup>, 2004 from the World Wide Web: <http://home.allgaeu.org/ndoell/work/ircpaper.htm>.
- Farfeleder, S. (2000). *The analysis of chat communication*. The website of English and American Studies at University of Salzburg: Austria. Retrieved June 2<sup>nd</sup>, 2004 from the World Wide Web: [www.sbg.ac.at/ang/projects/ps\\_s00/webfiles/farfeleder.htm](http://www.sbg.ac.at/ang/projects/ps_s00/webfiles/farfeleder.htm).
- Graddol, D. (1989). Some CMC discourse properties and their educational significance. In Robin, M. & Anthony, K. (Eds), *Mindweave: Communication, computers, and distance education* (pp.236-241). Oxford, Pergamon Press.
- Halliday, M.A.K. (1992). *Spoken and writing language* (3<sup>rd</sup> ed.). Oxford: Oxford University Press.
- Harris, L.D. & Wambeam, C.A. (1996). The internet-based composition classroom: A study in pedagogy. *Computers and Composition*, 13, 353-371.
- Hubbard, D. & Walberg, H.J. (1997). Student view of computer-composition effects on writing. *Computers and Composition*, 15, 59-71.
- Johanyak, M.F. (1997). Analyzing the amalgamated electric text: Bringing cognitive, social, and contextual factors of individual language users into CMC research. *Computers and Composition*, 14, 91-110.

Ablex Publishing Corporation.

- Kaplan, R.B. & Grabe, W. (2002). A modern history of written discourse analysis. *Journal of second language writing*, 11, 191-223, Pergamon.
- Lim, J.K. (2003). *Study of English patterns on Internet*, Master's thesis, Graduate school of education, Sungshin Women's University.
- McCarthy, M. (1991). *Discourse analysis for teachers*. Cambridge: Cambridge University Press.
- Park et al. (2001). *Dictionary of Applied Linguistics*. Seoul: Kyungjin Publisher.
- Park, K.J., Nakano, M. & Lee, H.K. (2003). *Cross-Cultural Distance Learning and Language Acquisition*. Seoul: Hankook Publishing Co.
- Stenström, A-B. (1994). *An Introduction to spoken interaction*. London and New York: Longman.
- Takayoshi, P. (1996). The shape of electronic writing: Evaluating and assessing computer-assisted writing processes and products. *Computers and Composition*, 13. 245-257.
- Ure, J. (1971). Lexical density and register differentiation. *Applications of linguistics: Selected papers of the second international congress of applied linguistics, Cambridge, 1969*, Cambridge: Cambridge University Press.
- Werry, C.C. (1996). Linguistic and interactional features of internet relay chat. In Herring. S. (Ed), *Computer-Mediated Communication: Linguistics. Social and Cross-Cultural Perspectives* (pp.13-28). Amsterdam: John Benjamins.