The effect of language complexity and group size on knowledge construction: Implications for online learning

Ines Khalsi Zaouchi
(PhD Candidate and Assistant Teacher, The Higher Institute of Studies Applied to Humanities, Tunisia)
ines.khalsi@gmail.com

(Received: 24.12.2012, Accepted: 19.02.2013)

Abstract
This study investigated the effect of language complexity and group size on knowledge construction in two online debates. Knowledge construction was assessed using Gunawardena *et al.*'s Interaction Analysis Model (1997). Language complexity was determined by dividing the number of unique words by total words. It refers to the lexical variation. The results showed that knowledge construction and group size are significantly and negatively correlated. Also, the study revealed that knowledge construction and language complexity are significantly and positively correlated. Furthermore, the study demonstrated that language complexity is a significant predictor of knowledge construction in online debates. Some actions should be undertaken to increase language complexity in order to foster knowledge construction in online debates.

Keywords: knowledge construction; group size; language complexity; online debates.

Introduction

As the Internet is increasingly growing, online education continues to grow too (Johnson & Aragon, 2003), a phenomenon expected to continue at a significant rate (Allen & Seaman, 2004). Online discussion forums, or Computer Mediated Discussions, are popular with educators who aim at using IT (Information Technology) to enhance the quality of learning. The use of computer-mediated-communication tools can present new ways to promote knowledge construction (Schellens & Valcke, 2006). Computer-mediated-communication tools can help make the construction of knowledge easier by working as a social medium to support students' learning by representing students' ideas and understandings in concrete forms (e.g., notes) so that ideas can be more developed via social interactions (e.g., questioning, clarifying) (Van Drie, Van Boxtel, Jaspers & Kanselaar, 2005). One example of such tools is the asynchronous discussion forum. The technology which is available in asynchronous online discussions provides a number of ways to foster the construction of collaborative knowledge, while asynchronicity offers learners the opportunity to interact at any time from any place (Scardamalia & Bereiter, 1994). The debate could be described as a constructive learning environment which offers multiple approaches and actual world examples of the topic of discussion, that encourages reflection, and that supports collaborative construction of knowledge via social negotiation (Jonassen, 1994).

Early analyses of computer-mediated communication using asynchronous tools tended to concentrate more on quantitative analysis of the data, especially on word counts and number of postings. Yet, although this method of analysis provides a survey of the interactions which occur online, it does not take into consideration the content of what is posted on the discussion boards. The analysis of the content of the discussion boards, thus, moves towards a more semantic labeling of propositions (Donnelly & Fitzpatrick, 2010). The assessment of co-construction of knowledge based on quantitative analysis of postings underestimates the complexity of the available issue. Although a quantitative analysis allows the researchers to understand some linguistic online behaviors, it does not allow deep investigation of the language complexity in order to pinpoint the collaborative learning among learners. Thus, linguistic models for a qualitative analysis of online discourses have been elaborated by several researchers; for example, Interaction Analysis Model by Gunawardena, Lowe and Anderson (1997).

More recently, some researchers have examined if group size might influence the levels of knowledge construction in online discussion forums. Schellens and Valcke (2006), for example, found that discussion in groups of about 10 participants resulted in larger proportions of advanced levels of knowledge construction. Hew and Cheung (2010) examined if there was any relationship between the frequency of advanced level knowledge construction occurrences and group size. The researchers found a significant positive correlation between the discussion group size and the frequency of advanced level knowledge construction occurrences. However, no indication was provided by Hew and Cheung (2010) about the optimal group size.

In fact, no research study investigating the impact of language complexity on knowledge construction in online conversations has been reported. Language complexity refers to the lexical variation of a given text. Consequently, this study makes an endeavour to provide some evidence that seems to be urgently needed. This paper addresses the effect of group size and language complexity on knowledge construction in online debates and tries to ask these two research questions: Is there a significant relationship between knowledge construction and group size in online debates? And, is there a significant relationship between language complexity and knowledge construction in online debates?

The study

The goal of this study was to build on the current literature through exploration of how group size impacts participants' construction of knowledge within a primary asynchronous environment. It also tries to investigate the impact of language complexity on knowledge construction. This study is a longitudinal case study because the data source is bounded by time and environment (Creswell, 1998).

Variables of the study

Knowledge construction

Knowledge construction refers to phases of interaction in the online debates. Phases of interaction were identified using Gunawardena et al.'s (1997) Interaction Analysis Model.

Group size

Group size of an online debate refers to the number of participants who were involved in the conversations. Two main forms of participation are identified in an online discussion environment: writing and reading (Hewitt & Brent, 2007). In this research paper, the focus is on the writing form of participation because writing is closely linked to discussion, and it is of greater importance than reading (e.g., when the student is answering postings from an existing discussion thread) (Guzdial & Turns, 2000). Moreover, writing is a more observable phenomenon than reading. In Debate A, group size is equal to 326 whereas in Debate B, group size is equal to 118.

Language complexity

Language complexity (LC) variable is determined by type token ratio (TTR), which is a measure of vocabulary variation within a written text or a person's speech. The type-token ratio has been shown to be a helpful measure of lexical variety within a text. The number of words in a text is often referred to as the number of tokens. However, several of these tokens are usually repeated. As long as there is only one type of word, the relationship between the number of types and the number of tokens is known as the type token ratio (TTR) (Williamson, 2009).

A high TTR indicates a large amount of lexical variation and a low TTR indicates relatively little lexical variation (Williamson, 2009). The following table features the different TTR levels:

Table 1: TTR level classifications

Level	Very low %	Low %	Average %	High %	Very high %
TTR	0-20	20-40	40-60	60-80	80-100

Informants

Informants of the study are 444 online debaters participating in two online debates:

A: 326 debaters participating in the online debate: "Technology in Education" retrieved from:

http://www.economist.com/debate/days/view/244,

B: 118 debaters participating in the online debate "Internet Democracy" retrieved from:

http://www.economist.com/debate/days/view/662.

Online debates sampling

The first online debate is entitled "Technology in education" and was retrieved from the website "The economist.com" on March 18th, 2011. It was carried over 11 days from the 15th till the 26th of October 2010 and comprised 371 comments. It was coded Debate A. The second online debate is entitled "Internet Democracy" and was also retrieved from the website "The economist.com" on April 13th, 2011. It was carried over 10 days from the 23rd February 2010 till the 4th February 2010 and comprised 128 comments. It was coded Debate B.

Interaction Analysis Model

The informants' online transcripts were analyzed qualitatively using Gunawardena et al. (1997) Interaction Analysis Model (IAM). The analysis is based on the five phases of knowledge co-construction that usually occur during online debates.

Gunawardena et al. (1997) stated that postings coded Level I and II "represent the lower mental functions", while postings coded level III, IV, and V represent the higher mental functions:

a) Level I – making statement of observation or opinion, statement of agreement among participants;

b) Level II - identifying areas of disagreement, asking, or answering questions to clarify disagreement;

c) Level III - negotiating the meaning of terms, ideas/co-construction of knowledge;

d) Level IV - testing of proposed synthesis or construction against existing literature or personal understandings, experiences; and

e) Level V - summarizing agreement/statements that show new knowledge construction, application of newly constructed ideas. In this study, we defined advanced levels of knowledge construction as levels II, III, IV, or V of the model.

Procedure

To apply the Interaction Analysis Model, I read the postings in the original sequence and assigned them one or more phases from the IAM. It is possible to code multiple sentences or a paragraph or two with a single phase; this is consistent with the original application of the IAM (Gunawardena et al., 1997). I calculated the frequencies of the coded phases for each posting and for each informant. Two raters, myself and an English assistant colleague, coded the online transcripts. In order to conduct inter-reliability checks, I used the most advanced phase from each posting as the basis for inter-rater checks (Beaudrie, 2000). Inter-rater differences were addressed following Chi (1997).

Postings were coded using the five phases of Gunawardena et al. (1997). For statistical correlation, Phase I was coded 1, phase II was coded 2, phase III was coded 3, phase IV was coded 4 and phase V was coded 5. The 'absence of phase' was coded 0. A second researcher reviewed the coding of the total postings in debate A and B. The inter-rater was selected based on her field of specialization, applied linguistics, and her familiarity with discourse analysis. The inter-rating training consisted of an independent review of the Interaction Analysis model. Her task was to review the coding made by the investigator. It was easy to achieve an agreement of 100% because coding disagreement concerned only 3 postings in Debate B. Total agreement was achieved after discussing discrepancies.

"TextMaster" was downloaded from the Internet. "TextMaster" is a software tool for rapid analysis and processing of fixed-length files. This software counts the number of tokens and types. Each posting was copied and entered in "TextMaster" to obtain the number of tokens and types. TTR was then processed for each posting through dividing the number of types by the number of tokens. The value obtained is referred to as language complexity. The mean language complexity was processed for the participants who sent two postings or more. Numerical data of TTR was turned into categorical data in order to process statistical analyses. Values belonging to the very low TTR levels were coded 1. Values belonging to the low TTR levels were coded 2. Values belonging to the average TTR levels were coded 3. Values belonging to high TTR levels were coded 4. Values belonging to very high TTR levels were coded 5.

The study investigated two online debates. Debate A comprises 326 participants and Debate B comprises 118 participants. Group size in Debate A was coded 1 and group size in Debate B was coded 2. The statistical data analysis was based on descriptive and analytical statistics. Descriptive statistics were used to calculate means and percentages of the selected variables of the study which are language complexity, knowledge construction, and group size. Correlation analysis was used to describe the relationship between the different variables. Spearman's Rho correlations were computed between different variables - language complexity, group size and knowledge construction - to detect any relationship between them. Simple regression analyses were computed on dependent and independent variables to determine the significant predictors of knowledge construction. Multiple regressions analyses were computed on dependent and independent variables to confirm simple regression results. The data were computed using the statistical Package for the Social Sciences (SPSS) 17.

Findings

Table 2 reveals that in Debate A the relationship between language complexity and knowledge construction is positive and highly significant at the 0.01 level of significance. It also shows that in Debate B the relationship between language complexity and knowledge construction is positive and significant at the 0.05 level of significance. These results imply that the higher the language complexity is, the higher the knowledge construction would be.

Table 2: Spearman's Rho correlations between language complexity and knowledge construction in Debate A and Debate B

	Knowledge construction	
	Debate A	Debate B
Language complexity	0.18** .001	0.21* .01

**Correlation is significant at the 0.01 level
*Correlation is significant at the 0.05 level

Table 3 shows that the relationship between knowledge construction and group size is negative and highly significant at the 0.01 level implying that the less important group size is, the more important knowledge construction would be.

Table 3: Spearman's Rho correlations between group size and knowledge construction

	Knowledge construction
Group size	-1.67** .00

**Correlation is significant at the 0.01 level
*Correlation is significant at the 0.05 level

Table 4 shows that language complexity has given non-significant results in the regression equation for knowledge construction in Debate A. However, Table 5 reveals that language complexity is the most consistent predictor of the variation observed in knowledge construction in Debate B. It accounts for 5.1 % of the observed variation. The regression equation is significant as shown by the t-value and the F-ratio.

Table 4: Simple regression of language complexity for knowledge construction in Debate A

	Language complexity	R^2	R^2(adj)	F-ratio
Knowledge construction	0.04 (0.71) NS	0.2%	0.1%	0.51 NS

1. Bracketed figures are t values.
2. ** = p < .01
3. * = p < .05

NS = non-significant

Table 5: Simple regression of language complexity for knowledge construction in Debate B

	Language complexity	R^2	R^2(adj)	F-ratio
Knowledge construction	0.22 (2.5) *	5.1%	4.3%	6.26 *

1. Bracketed figures are t values.
2. ** = p < .01
3. * = p < .05

NS = non-significant

Table 6 reveals that group size gives non-significant results in the regression equation for knowledge construction. Consequently, group size is not a significant predictor of knowledge construction.

Table 6: Simple regression of group size for knowledge construction

	Group size	R^2	R^2(adj)	F-ratio
Knowledge construction	2.58 (-0.33) NS	3.1%	2.9%	14.24 NS

1. Bracketed figures are t values.
2. ** = $p < .01$
3. * = $p < .05$

NS = non-significant

Table 7 shows that when group size is added to language complexity in the same regression equation for knowledge construction, the adjusted R^2 falls from 4.3% to 2.8%. Since the t-value is not significant for the two variables, group size does not help knowledge construction. Consequently, the best regression fit is the simple regression of language complexity for knowledge construction in Debate B.

Table 7: Multiple regression of language complexity and group size for knowledge construction

	Language complexity	Group size	R^2	R^2 (adj)	F-ratio
Knowledge Construction	0.22 (2.5) *		5.1 %	4.3 %	6.26 *
	0.07 (0.8) NS	-0.34 (-0.035) NS	3.2 %	2.8 %	7.39 NS

1. Bracketed figures are t values.
2. ** = $p < .01$
3. * = $p < .05$

NS = non-significant

Discussion

In both debates the results show that language complexity and knowledge construction are significantly correlated. Correlation is positive and highly significant in Debate A and positive and significant in Debate B suggesting that an increase in language complexity generates an increase in knowledge construction. This finding implies that using rich and complex vocabulary results in consistent conversations which tend to engender various ideas, opinions and viewpoints. Consequently, this could promote negotiation and higher order thinking. Furthermore, findings show that language complexity is a significant predictor of knowledge construction. Thus, generating a high lexical variation may foster high levels of knowledge building. Therefore, educators should mainly focus on techniques that promote vocabulary richness.

Besides, students' participation may vary according to the mastery of the language used. Many learners may feel some difficulties when communicating in their second or foreign language which implies that asynchronous online environment may be an effective tool in evaluating the students' language proficiency. Furthermore, some actions should be undertaken to help learners enhance their language level such as undertaking reading and writing sessions. The stress should be placed on English, which is an international language. Participating in such debates using the second or foreign language would be an efficient practice. Online communication environments are empowering tools for non-native speakers.

In order to promote rich and consistent online conversations, students' online participation should be fostered. Different roles can be attributed to students. Some of them can play the role of moderators. They may be fight-flaming and stop altercation, though. Others should have the role of summarizers, summarizing long and frequent postings in order to facilitate the interaction. A group of participants may also find appropriate theories to back up informants' statements, thus playing the role of theoreticians. Giving such responsibilities to students will not only facilitate communication but will also stimulate them to participate actively in the discussion, promoting, therefore, language complexity and knowledge construction.

The results also revealed that the correlation between group size and knowledge construction is negative and highly significant. This implies that high levels of knowledge construction are achieved by informants participating in smaller forums. These findings contradict the ones reported in Hew and Cheung (2010). In fact, allowing for an ongoing increase in the discussion size may have several limitations. First, it may result in 'reading without writing' on the part of the participants. Second, large groups or conversations require huge cognitive efforts from the participants to react to others. This could result in reading boredom.

Hew and Cheung (2010) suggest a group size of about 10 participants in order to form a critical mass to lead the discussion to advanced levels of knowledge building (P.431). Students' group size should be limited in order to avoid learners' exhaustion and withdrawal from the debate. In fact, a big-sized group often results in a big-sized conversation; and students would be overwhelmed by the number of postings. Limiting students' number would therefore help them go through the five phases of interaction.

Limitations and future research

The main limitation of this paper is that it investigated only two online discussions. To obtain significant results on the effect of group size on knowledge construction in online debates, the number of forums should be increased. One of the main limitations of this type of research is the subjectivity of coding. The classification of messages is open to individual interpretation. Using Interaction Analysis Model is based mainly on personal opinion. The content might be understood differently by coders resulting in different phases of coding.

This research study could be undertaken in other contexts and by including other variables. For instance, it could be conducted in another medium of communication. Other factors that influence knowledge construction could be considered, such as the amount of participation. Further research is also needed to discover whether the type of knowledge or the amount of knowledge are significant predictors of participation level and knowledge construction that occur in online debates. It would also be quite interesting to study knowledge construction in online conversations from a sociolinguistic perspective and find out how social variables such as age, location, social status, time or Internet accessibility could be related to level of knowledge constructed but future data collection and analysis are required for more rigorous findings.

References

Allen, I., & Seaman, J. (2004). *Entering the mainstream: The quality and extent of online education in the United States, 2003 and 2004.*Wellesley, MA: Sloan Consortium

Beaudrie, B. P. (2000). *Analysis of group problem-solving tasks in a geometry course for teachers using computer mediated conferencing*. Unpublished doctoral thesis, Montana State University, Bozeman

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. The *Journal of the Learning Sciences*, *6*(3), 271–315.

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five designs*. Thousand Oaks, CA: Sage.

Donnelly, R., & Fitzpatrick, N. (2010). Do you see what I mean? Computer-mediated discourse analysis. In

Donnelly, R., Harvey, J., & O'Rourke, K. (Eds.) *Critical Design and Effective Tools for E-Learning in Higher Education: Theory into Practice*. Hershey, PA: Information Science Reference.

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a Global Online Debate and the Development of an Interaction Analysis Model for Examining Social construction of knowledge in Computer Conferencing. *Educational Computing Research*, *17*, 397-431.

Guzdial, M., & Turns, J. (2000). Effective discussion through a computer-mediated anchored forum. *Journal of the Learning Sciences*, *9*(4), 437-469.

Hew, K. F., & Cheung, W. S. (2010). Higher-level knowledge construction in asynchronous online discussions: An analysis of group size, duration of online discussion, and student facilitation techniques. *Instructional Science*. DOI: 10.1007/s11251-010-9129-2

Hewitt, J., & Brett, C. (2007). The relationship between class size and online activity patterns in asynchronous computer conferencing environments. *Computers and Education*, *49*(4), 1258-1271.

Johnson, S. D., & Aragon, S. R. (2003). An instructional strategy framework for online learning environments. *New Directions for Adult and Continuing Education*, *100*, 31-43.

Jonassen, D.H. (1994).Thinking Technology: toward a constructivist design model. *Educational Technology*, *4*, 34-37.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences*, *3*(3), 265-283.

Schellens, T. & Valcke, M. (2006). Fostering knowledge construction in university students through asynchronous discussion groups. *Computers & Education*, *46*(4), 349-370.

Van Drie, J., Van Boxtel, C., Jaspers, J., & Kanselaar, G. (2005). Effects of representational guidance on domain specific reasoning in CSCL. *Computers in Human Behavior*, *21*(4), 575–602.

Williamson, G. (2009). *Type-Token Ratio*. Retrieved from :http://www.speech-therapy-informationand resources.com/type-tokenratio.html. Accessed 31.01.2010.