RESEARCH ARTICLE

AN EMPIRICAL STUDY OF ARGUMENTATION BY QUALITATIVE CODING AND ONLINE CORPUS ANALYSIS TOOLS

Huangfu Wei
School of Foreign Languages,
North China Electric Power University, Beijing, China
Email:mailtohf@163.com

Abstract

Argumentation poses challenges for EFL students in the sense of free-flowing discourse. Due to the nature of online discussion groups as self-organized, informal circumstance, investigating argumentation in such environment offers an opportunity for obtaining insight into how people debate with one another online. In this empirical study, both Atlas-ti qualitative coding tool and WMatrix online corpus analysis tool are used in analyzing the distribution and co-occurrences of discoursal function codes and linguistic forms. Then, the detailed analyses of the gathered data shed light on patterns of co-occurrence between linguistic categories and discoursal function codes to illustrate online argumentation. This research highlights the necessity of incorporating this combined qualitative and quantitative methodology in discourse analysis.

Keywords: argumentation; qualitative coding; ANOVA analysis; Atlas-ti; WMatrix

1. Introduction

Studies foci on argumentation have been changed from the perspective of logic and philosophy to the contemporary perspectives of informal logic and pragma-dialectics. According to Walton (2006), to argue is to give reasons to support a proposition so as to persuade others or get rid of the criticisms. According to van Eemeren and Grootendorst (2004), arguing with others is to justify or refute the proposition's acceptability in the standpoint. So, claims and challenges stand in the prominent position in those definitions of argumentation. Similar to other face-to-face talk, it is common to note in argumentation research that argumentational discoursal functions are not always easily identified (Erduran et al., 2004). In this study, the emphasis is on the looser perspective of online

discussions instead of the efforts to succeed or oppose in talks because a clear and stable position is not always possible in a free exchange of opinions. Potential argumentational moves together with those that are unambiguously argumentational will all be incorporated in such an online context.

The quality of students' argumentation during asynchronous online discussions (Nussbaum et al. 2007) and on-line synchronous argumentation learning (Yeh et al., 2010) have been well explained. Online environment is an ideal place for informal argumentation because the virtual cyberplace offers opportunities for participators to claim and challenge and assess and explain in discourses on a given topic, and also for team-members to cooperate in these debates. Smooth and free exchange of opinions is most likely to happen in the

online teaching-learning paradigm for its collaborative features. Learners construct and balance arguments and counter-arguments in order to find resolutions and collaborations rather than confrontations (Weinberger and Fischer, 2006).

With new and authentic contexts, the asynchronous feature of many such online discussions not only provides more flexibility, such as "wait time", but also affords participation from a heterogeneous population of students. Participants can use a variety of discoursal functions and semantic forms in their online exchanges with others. Thus, it is generally agreed that some computer-supported scaffolds can construct different components of arguments, such as claims (Kenyon and Reiser, 2006), challenges (Schwarz and Glassner, 2003), or some specific discursive moves, etc.

Very few instances could be similar to such an opinion exchange place as the web for its adequate chances in informal debate on an equal status. Thus, researching the discourse of online argumentation offers an opportunity for understanding how people debate with each other in such an informal circumstance.

2. Methodology

In this study, EFL students who major in ten subjects, of which five groups are students of science specialty and other five groups are students of art specialty, were recruited to participate in online discussions of texts they learned in an extensive class across the semester. They registered in separate teaching blogs created by the researcher on the website sponsored by the university to provide easier access for the participants to perform the discussions and for the researcher to moderate the discussions and collect data afterwards. Subjects were chosen on diverse factors of gender, major, language proficiency and computer skills to make the subject population reasonably representative of a much larger population so as to achieve greater validity. A combination of two tools is chosen for this study, i.e. a manual qualitative coding software tool, i.e. Atlas-ti and an automated quantitative tool of corpus linguistics i.e. WMatrix (Rayson, 2008). So, qualitative and quantitative analysis of linguistic

forms and discoursal functions in argumentation could be made.

Atlas-ti software can be used to highlight areas of text and code them with a specific name. In Atlas-ti software, such highlighted texts are referred as quotations. In this study, texts of different discoursal functions were highlighted as quotations and coded with corresponding names. Different quotations on different codes can be extracted to build specific corpora. These were further investigated by using WMatrix, an online automated quantitative corpus analysis tool. A two-level of codes was devised, i.e. phase codes and discoursal function codes. Engagements in close discussions on the chosen topic were coded as on-discussion phase, otherwise as off-discussion phase where the subjects talk off-track and which were of less concern of this study. Two types of discourse function codes were agreed upon, i.e. discussion codes, such as "claim", "challenge" and "co-construction" and thematic codes, such as "assessment" and "explanation". These two codes can co-occur when participants are engaged in argumentation over a topic.

Corpora have been widely used in a variety of researches. Automated software for researching a corpus facilitates quantitative investigation. But, their uses for examining argumentation are rare cases (Coffin and O'Halloran, 2009). Through WMatrix, this research made quantitative linguistic study of discoursal function corpora which were built from corresponding quotations in Atlas-ti for discoursal function codes. Then, keyness values of loglikelihood were calculated for these linguistic categories.

3. Investigating Discourse Codes via Atlas-ti

Atlas-ti can not only be used to code the discourses qualitatively, but also can provide quantitative data on the use frequency of those codes and the word counts of all quotations for every code. Using the analysis of two-way repeated measure ANOVA, the overall discussion code uses for the students of science and students of art and detailed analysis of the three different types of discussion codes are provided in table 1. As to the overall discussion code uses, there was both

significant difference across science and art students, and also significant difference between the three discussion code uses. Overall, since the interaction of these two factors was 0.42, which was larger than 0.05, it indicated the two factors were independent of each other and no significant interaction effect was observed across the groups.

Table 1: Two-way ANOVA results for types of specialties and discussion codes

	Type III sum of squares	df	Mean squares	F	Sig.
Specialty	300.833	1	300.833	1.69	0.000***
Discussion	2000.6	2	1000.3	12.063	0.001**
Specialty *discussion	643.267	2	321.633	3.879	0.42
Error	1326.8	16	82.925		

Note: *p<0.05, **p<0.01, ***p<0.001

Using the Atlas-ti's query tool, co-occurrences of codes can be established to what is the relationship between thematic codes (assessment/explanation) and discussion codes (claim/challenge/co-construction). It is found out that 143 claims co-occur with assessment code and 76 claims with interpretation code out of the 299 claims in total. This shows claims tend to be made to assess rather than to explain. Similarly, out of the 144 challenges, they are more likely to co-occur with assessment than explanation, 61 and 43 co-occurrences respectively. However, a much stronger tendency is observed towards explanation than assessment in co-construction. As is shown in table 2, there is no significant difference for co-occurrences of discussion codes and thematic codes.

Table 2: One-way ANOVA results for co-occurrences of discussion and thematic codes

	Assessment	Explanation	F	Sig.
Claim	143	76	0.122	0.744
Challenge	61	43		
Co-construction	32	80		

According to types of discussion codes and thematic codes, quotations on these codes were extracted, and converted into text files, thus three discoursal function-based corpora and two thematic-based corpora were compiled. To enhance the quantitative values of this research and show the extent of quotations across the corpora, word counts are also gained in Atlas-ti for all quotations attached to a type of code. As is shown in table 3, there is a significant difference for code and word count. While there are higher frequency of occurrences of "claim" and "challenge" codes, co-construction make up a larger body of texts. Similarly, with thematic codes, while there is more assessment than explanation codes, there are more words in explanation than in assessment.

Table 3: Results gained via one-way ANOVA for code and word count across the corpora

.	Word count	Code count	F	Sig.
Claim	10003	299	11.362	0.01*
Challenge	2948	144		
Co-construction	14640	112		
Assessment	23901	121		
Explanation	29516	111		

Note: *p<0.05, **p<0.01, ***p<0.001

4. Investigating Linguistic Categories via WMatrix

Linguistic categories with statistically salient values need also to be inspected in the corpora generated via Atlas-ti to judge their qualitative significance. The POS tags and semantic domains that WMatrix uses are extensive and

therefore, to explore the above qualitative linguistic categories, the researcher needs to be selective. In order to produce a systematic comparison of effusiveness and its intensity in the three corpora, this study explores through the semantic domain tags "interested/excited/energetic" (which tags lexis such as "interested/ing", "excited/ing", "gripped/ing", "engaged/ing") and the "degree adverb' POS tag which tags lexis such as "so" and "very". Also, this study compares in WMatrix the semantic domain "exclusivisers /particularisers", which contains the single word adverb, "just". Present and past tense of lexical verbs are explored in the three corpora using POS tags for these tenses; the pronouns "I" and "he"/"she" are explored via corresponding POS tags. In sum, with these foci, insight can be gained into some significant semantic dimensions of the discourse.

WMatrix has British National Corpus (BNC) Sampler available for this research as the reference corpus. BNC sampler corpus in WMatrix contains both written English and spoken English. In this research, the BNC sampler spoken corpus is chosen for comparing data with the three corpora, claim, challenge, co-construction, derived from Atlas-ti qualitative coding of data to get the keyness of loglikelihood value for the aforementioned specific semantic features of the online argumentation. When indicating quantitative results, much of this is done via brackets containing two figures in the table 4: the first refers to frequency and the second refers to log likelihood (LL) value at ≥7 (i.e. measure of statistical significance or keyness) proposed in Baker (2004).

Table 4: Results of keyness of semantic features across the three corpora

	Claim	Challenge	Co-construction
1	521; LL=27.3	205; LL=35.3	451; LL=271.7
2	318; LL=88.7	82; LL=31.6	263; LL=11.4
3	Null	Null	143; LL=22.2
4	47; LL=89.6	Null	Null
5	221; LL=210	Null	232; LL=15.5
6	97; LL=10.4	Null	Null

Note:1=I/he/she; 2=past tense of lexical verbs; 3=present tense for lexical verbs; 4= interested/excited/engergetic; 5=degree adverbs; 6=exclusivisers/particularisers; Null=no keyness or log likelihood＜7; the first number is frequency and LL represents log likelihood

From table 4, claims are mostly made in the past tense than in the present tense. The opposite is true of co-constructions, which have keyness in both past and present tenses but lower keyness than those in claim and challenge corpora. In another words, in co-construction, participants are less likely to be reporting pre-formulated individual opinions than in claims. This may be clearly seen from the quantitative perspective of analysis via Atlas-ti that claims predominantly co-occur with assessment, while co-constructions mostly co-occur with explanation, which also explains why keyness in the co-construction corpus is relatively higher for the pronouns compared with the claim and challenge corpora. In short, no individual speaker can dominate such explanations. In the co-construction, the semantic domain of exclusivisers/particularisers

do not have keyness, and in both co-construction and challenge corpora the degree adverbs do not have keyness, while the log likelihood values for those in the claim corpus are much larger and show keyness. This might be because claims together with assessment are more holistic, more complete than explanatory co-constructions, and that challenges are expressed with less mitigation or hesitancy than claims. Moreover, the fact that "interested/excited/engergetic" and degree adverbs have keyness in the claim corpus but not in the challenge corpus seems to indicate that challenges are expressed with less effusion than claims.

5. Conclusion

Online argumentation discourse is characterized by appropriate arrangement of

time/tense, pronouns, hesitancy/mitigation and effusiveness. The results show that the participants do not come to the discussions with stable, well-thought defensible positions. Their discussions seem not only to retrieve those bits from memory, but also to think in co-constructive discourse with other participants. Also, forming explanatory ideas in that debate is more co-constructed in the present than in the past. This collaborative discourse in online argumentation consists of free-flowing discussions with re-directive challenges to energize co-construction. In sum, this study has tentative evidences to show how people argue with each other online. However, more evidences should be gathered to further explore this across a larger population with more scientifically mixed designs of the research. Finally, the significance of using corpus tools and quantitative coding tools in discourse analysis has been fully demonstrated in this study.

References

[1]. Walton, D., *Fundamentals of Critical Argumentation*, Cambridge University Press, 2006.

[2]. van Eemeren, F. and R. Grootendorst., *A Systematic Theory of Argumentation: The Pragma-Dialectical Approach*. Cambridge University Press, 2004.

[3]. Erduran, S., S. Simon, and J. Osborne.,Tapping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse,*Science Education*, 2004, vol. 88, pp. 915–33.

[4]. E. M. Nussbaum, D. L. Winsor,Y.M. Aqui and A.M. Poliquin, Putting the pieces together: Online argumentation VEE diagrams enhance thinking during discussions,*Computer-Supported Collaborative Learning*, 2007, vol. 2, pp. 479–500.

[5]. Kuan-Hue Yeh and Hsiao-Ching She, On-line synchronous scientific argumentation learning: Nurturing students' argumentation ability and conceptual change in science context,*Computers & Education*, 2010, vol. 55, pp. 586-602.

[6]. Weinberger, A. and F. Fischer.,A frame work to analyze argumentative knowledge construction in computer-supported collaborative learning, *Computers and Education*, 2006, vol. 46, pp. 71–95.

[7]. Kenyon, L., & Reiser, B. J. A functional approach to nature of science: Using epistemological understandings to construct and evaluate evidence, Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, 2006.

[8]. Schwarz, B. B., & Glassner, A., The blind and the paralytic: Supporting argumentation in everyday and scientific issues. In J. Andriessen, M. Baker, & D. Suthers (Eds.), *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments*, Norwell, MA:Kluwer, 2006, pp.227–260.

[9]. Rayson, P., Wmatrix:a web-based corpus processing environment, Computing Department, Lancaster University, 2008, Available at: http://ucrel.lancs.ac.uk/wmatrix.

[10]. Baker, P.,Querying key words: questions of difference, frequency, and sense in key words analysis,*Journal of English Linguistics*, 2004, vol.32, pp. 346–59.