

Learning Analytics To Identify Exploratory Dialogue within Synchronous Text Chat

Rebecca Ferguson and Simon Buckingham Shum

SocialLearn Project, Knowledge Media Institute, The Open University
Milton Keynes, MK7 6AA, UK
r.m.ferguson/s.buckingham.shum@open.ac.uk

Abstract. While generic web analytics tend to focus on easily harvested quantitative data, Learning Analytics will often seek qualitative understanding of the context and meaning of this information. This is critical in the case of dialogue, which may be employed to share knowledge and jointly construct understandings, but which also involves many superficial exchanges. Previous studies have validated a particular pattern of “exploratory dialogue” in learning environments to signify sharing, challenge, evaluation and careful consideration by participants. This study investigates the use of sociocultural discourse analysis to analyse synchronous text chat during an online conference. Key words and phrases indicative of exploratory dialogue were identified in these exchanges, and peaks of exploratory dialogue were associated with periods set aside for discussion and keynote speakers. Fewer individuals posted at these times, but meaningful discussion outweighed trivial exchanges. If further analysis confirms the validity of these markers as learning analytics, they could be used by recommendation engines to support learners and teachers in locating dialogue exchanges where deeper learning appears to be taking place.

Keywords. educational dialogue, text chat, instant messaging, exploratory dialogue, learning analytics, synchronous dialogue

Introduction

Learning resources are being uploaded to the Internet at such a rate that is increasingly likely that individuals will find themselves adrift in an ‘ocean of information’ [1, p136]. In summer 2010, Apple reported that iTunesU contains more than 350,000 audio and video files [2], and this vast online storehouse is dwarfed by YouTube, where 20 hours of video are uploaded to the site every minute [3]. At the same time, universities are increasingly sharing online seminars and conferences that stretch over hours or days. These resources that extend over time are difficult to scan or assess quickly and so learners and teachers must rely on basic, often misleading, cues such as title, keyword and producer when deciding whether to make use of a resource. Analytics are therefore needed that can distinguish between resources that extend over time, identifying those that support learning.

Although these resources are not necessarily text-based, many have series of text associated with them, in the form of comment streams or chat. In other contexts, various approaches have been used to identify and classify forms of learning dialogue and academic dialogue but these are typically dependent on the use of grammatically correct, carefully punctuated and formally structured text [4,5]. Synchronous textual dialogue is likely to be more akin to speech than to formally constructed prose [6]. It is therefore relevant to look at how people build knowledge together through speech. In face-to-face settings, Mercer and his colleagues [7-11] have distinguished three social modes of thinking used by groups of learners: disputational, cumulative and exploratory. Of the three, exploratory dialogue is the type considered most educationally desirable by teachers [12]. Mercer and Littleton [10, 62] provide a clear description of its use in a school environment:

Exploratory talk represents a joint, coordinated form of co-reasoning in language, with speakers sharing knowledge, challenging ideas, evaluating evidence and considering options in a reasoned and equitable way. The children present their ideas as clearly and as explicitly as necessary for them to become shared and jointly analysed and evaluated. Possible explanations are compared and joint decisions reached. By incorporating both constructive conflict and the open sharing of ideas, exploratory talk constitutes the more visible pursuit of rational consensus through conversation.

Exploratory dialogue is a form of discourse that may be found in both online and offline learning environments [6,13], where it can be taken as an indication that learning is taking place and that learners are going beyond a simple accumulation of ideas. The research reported here therefore asks: *Could the identification of exploratory dialogue within the synchronous textual chat associated with online resources help to identify resources and sections of resources that support learning?*

Data collection and preparation

In order to investigate these questions, data were collected from Elluminate®, a web conferencing tool that supports chat alongside video, slides and presentations. The focus was on synchronous discussion related to an online teaching and learning conference targeted at higher education researchers and practitioners. The Elluminate text chat in four conference sessions, each between 150 and 180 minutes in length (24,530 words in total) was investigated. The conference timetable was used to subdivide these four conference sessions into smaller units, including pre-session chat, post-session chat, conference introduction, groups of short talks, longer talks, moderated discussion and keynotes.

The four conference sessions were all archived and made public by the organizers. Sociocultural discourse analysis [14] was used to identify words that could be indicative of exploratory dialogue. These included:

- **Challenges** eg But if, have to respond, my view
- **Critiques** eg However, I'm not sure, maybe
- **Discussion of resources** eg Have you read, more links

- **Evaluations** eg Good example, good point
- **Explanations** eg Means that, our goals
- **Explicit reasoning** eg Next step, relates to, that's why
- **Justifications** eg I mean, we learned, we observed
- **Others' perspectives** eg Agree, here is another, take your point

Ninety-four words and phrases were identified in this way. Some words, phrases and punctuation, which initially appeared to be good indicators, were discarded because they were often used for finding out more about the conference, its tools and participants, rather than its content. For example, interrogatives and question marks were often associated with comments such as 'Can you still hear?' or 'what's everyone doing for coffee???' Once exploratory markers had been identified, the Elluminate chat was pasted into Microsoft Word, where a simple 'find and replace' Apple Script program was used to highlight the key words and phrases. The data was then transferred to Excel for more detailed analysis. Table 1 gives an example of data coded as exploratory.

Table 1: Dialogue coded as exploratory (real names removed). Each row of the table represents one contribution. Words in bold have been highlighted by the analysis.

what about quality control? Not all authors are as skilled as [named individual]...
He could get some dreadful blunders in public!
If you're doing it in public and people are following, that wil help in terms of quality control for egregious errors. However , if you're writing it in a vacuum, then that's not going to work so well, I guess.
Also , not everyone's blog has the traffic [named individual]'s has
Is it only skills that are needed, is it? There's something to do with attitude (to criticism, to mistakes, etc) – [named individual]? (sorry, no mic)
Shouldn't, in theory, course authors be writing carefully in the first place, because it's going to be seen by hundreds or thousands of students?
I would just note that a course team of 6 or 7 people can feel plenty public enough when you are trying to form thoughts - this is back to my point about there being stages where its good to be quite closed while you evolve an idea and approach
@[other conference participant] - yes, I think that educators will get a lot more out of this scary paradigm shift if they have set up some peers who also want to learn
[Another conference participant] Definitely or even interested "amateurs".

Once key words had been highlighted, the postings were divided according to the timings on the official conference timetable, and the use of exploratory dialogue in each section was calculated. As postings are short and clearly delineated, the posting was taken as the unit of analysis, and so an entire posting containing one or more markers of exploratory dialogue would be coded as exploratory.

The conference included two morning sessions and two afternoon sessions. For clarity, the analysis described here focuses on one afternoon, which preliminary analysis highlighted as containing the greatest concentration of exploratory markers. During that afternoon, the Elluminate session was divided into four sections: a set of

short talks, moderated discussion, keynote, and then chat between the scheduled end and the actual close of the Elluminate session.

Table 2: Analysis of contributions to the synchronous Elluminate text chat during one continuous afternoon conference session.

	Short talks (60 mins)	Discussion (45 mins)	Keynote (45 mins)	Chat (8 mins)
<i>Posts per minute</i>				
Mean number of posts per minute	2.4	4.6	5.8	8.8
Mean wordcount per minute	21.3	47.8	67.5	57.6
Mean number of exploratory posts per minute	0.4	1.2	1.6	0.9
Mean word count of exploratory posts per minute	5.2	18.6	31.1	11.8
<i>Posts per contributor</i>				
Mean number of posts per contributor	5.4	5.6	10.1	3.3
Mean word count per contributor	47.3	58.2	117.7	22.0
Mean exploratory posts per contributor	0.9	1.5	2.8	0.3
Mean word count of exploratory posts per contributor	11.4	22.6	54.2	4.5

Table 2 presents a summary of the analysis. As the length of the sessions ranged from 8 to 60 minutes, contributions were first classified by time. This showed that the most posts per minute took place during the informal chat session, whereas the most exploratory posts were contributed while the keynote was in progress. The series of short talks at the beginning of the afternoon appeared to be associated with the lowest levels of talk, whether exploratory or not.

The data were then classified by contributor (only participants who made some contribution to the live chat were included within these figures). Once again, a large amount of exploratory activity was evident during the keynote, whereas the many contributions during the informal chat were found to be short and lacking in exploratory talk. When analysed in this way, contributors were seen to be contributing longer, more thoughtful posts during the short talks at the beginning of the afternoon but, once again, the exploratory dialogue was less evident during these short talks than during the moderated discussion or the keynote.

As Elluminate identifies who has posted each comment in the text chat, it was also possible to consider the postings of individuals. Table 3 summarises analysis of the

third session in the afternoon, the keynote talk, which earlier analysis had shown to be the session associated with the most exploratory dialogue. The table compares the contributions of the five people who posted the most contributions during this session. The moderator (M) was very active during this session, posting 32 times, including long and exploratory posts that accounted for 42% of their total wordcount.

For all these individuals who posted a large number of posts, more than a fifth of their words were in contributions containing exploratory markers. However, there are notable differences within these groups, and C stands out as a high-volume poster with 75% of her total words in posts containing exploratory markers. These figures were typical of those in other sessions – the moderator was consistently one of the most active contributors. Although individuals’ interest and attention clearly fluctuated according to session, C was often among those with the highest percentage of exploratory posts in a session.

Table 3: Analysis of the contributions of the five individuals who contributed the most posts during the keynote session.

Contributor	A	B	C	D	M
Posts (Mean = 5.6)	16	16	17	27	32
Wordcount (Mean = 58)	111	183	297	210	253
Exploratory posts (Mean = 1.5)	3	3	11	6	7
Exploratory wordcount (Mean = 22.6)	43	38	224	75	105
Exploratory posts as % of personal posts	19%	19%	65%	22%	22%
Exploratory wordcount as % of personal wordcount	39%	21%	75%	36%	42%

Discussion

Preliminary analysis of the data suggests that markers of exploratory dialogue can be used to distinguish meaningfully between Elluminate sessions and to support evaluation of those sessions. The markers proved to be a more nuanced tool than generic analytics, such as simply counting the numbers signed in for an Elluminate session, or contributing to the text chat. Peaks of posting activity were associated with the end of Elluminate sessions, when many participants were thanking speakers and saying goodbye, while others were discussing what they had learned. Peaks of exploratory activity, on the other hand, were associated with periods set aside for discussion and keynote speakers. Fewer individuals posted at these times, but meaningful discussion outweighed trivial exchanges.

Exploratory markers indicate the importance of context when assessing learning dialogue. When several speakers were presenting in close succession, posting activity was relatively low, but increased as the presentations came to an end. However, when speakers had time to engage in discussion as part of their allotted timeslot – as was the case with the keynote speaker – meaningful exchanges peaked. Unscheduled chat

at the beginning of Elluminate sessions tended to be primarily social in nature, while unscheduled chat was likely to include many more exploratory exchanges.

This has implications for those scheduling online conferences – clearly-flagged discussion sessions related to presentations will be easier to find in the archives than discussions that overrun into other sessions. Discussion continues after scheduled sessions, so it could prove useful to leave Elluminate sessions open for chat for some time after the end of the scheduled presentation.

Not all exploratory dialogue related to conference content – there was considerable discussion of online conferences and of social issues. A future set of exploratory markers should identify keywords such as ‘mike’, ‘sound’ and ‘how are you’ that would signal a move away from discussion of content. At this stage, though, analysis suggests that time needs to be set aside for these exchanges, to avoid distraction or cognitive overload when presentations begin.

Areas for further investigation

Data analysis covered two complete days of online conference, and only a representative sample can be presented in a paper of this length. However, the analysis to date is clearly limited in its scope and there is a pressing need for evaluation of the reliability and validity of these presumed markers of exploratory dialogue – both individually, and as a set. If this set, or an amended set, of markers can be shown to be reliable and valid it will be important to attend to both context and practicalities. Exploratory dialogue is not necessarily focused on learning about content – individuals and groups are also likely to be learning about the tools they use (such as Elluminate) and the people with whom they are interacting. This learning dialogue is of less interest for people participating after the event, as they are neither using the same tools in the same way nor interacting with the same people. From a practical perspective, the current analysis is mainly carried out manually and in future it will be necessary to investigate how this process can be carried out automatically in order to benefit both learners and educators.

Compared to other computational linguistics approaches to text analysis, the approach presented in this paper is very simple; we are testing the limits of the simple exploratory dialogue markers described. In parallel, however, we are also beginning to test more complex forms of computational rhetorical analysis as described by Sándor [15,4], as a way to detect linguistic phenomena associated with the making of knowledge level claims around open educational resources, on which we hope to report in future work.

Conclusion

Although the conference sessions studied here are freely available as open online resources, they are both difficult and time-consuming for users to navigate. The published timetable of the conference gives some guidance, but is limited because a few sessions were reorganized, started late or overran. Some provoked little debate, whereas others inspired discussion which extended far beyond the scheduled time

period. The conference also included set-up sessions and breaks, during which talk turned to the practicalities of microphone use, and the absence of virtual biscuits. There is therefore a need for analytics that will allow learners to locate sections of an Elluminate session that clearly support learning.

At the same time, both learners and educators can benefit from tools that allow them to use Elluminate and other, similar, resources more effectively. Analytics can be used to distinguish different types of contribution to text chat, and to support learners who wish to engage in more fruitful learning discussion. They can be used to help educators schedule events in order to support discussion, and to model exploratory dialogue within that discussion.

References

1. Roach SS (1988) Technology and the services sector: America's hidden competitive challenge. In: Guile BR, Quinn JB (eds) *Technology in Services: Policies for Growth, Trade and Employment*. National Academies Press, Washington DC, pp 118-139
2. iTunes U downloads top 300 million. Apple press release (24 August 2010) (2010) <http://www.apple.com/pr/library/2010/08/24itunes.html>.
3. Zoinks! 20 hours of video uploaded every minute! YouTube blog (20 May 2009) (2009) <http://youtube-global.blogspot.com/2009/05/zoinks-20-hours-of-video-uploaded-every-20.html>.
4. Sándor Á, Vorndran A (2009) Detecting key sentences for automatic assistance in peer review research articles in educational sciences. Paper presented at the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries, ACL-IJCNLP 2009, Suntec, Singapore,
5. Whitelock D, Watt S (2007) Open Mentor: supporting tutors with their feedback to students. Paper presented at the 11th CAA International Computer Assisted Assessment Conference, 10/11 July 2007, Loughborough, UK,
6. Ferguson R (2009) *The Construction of Shared Knowledge through Asynchronous Dialogue*. PhD, The Open University, Milton Keynes. <http://oro.open.ac.uk/19908/>
7. Mercer N (1995) *The Guided Construction of Knowledge: Talk amongst Teachers and Learners*. Multilingual Matters Ltd, Clevedon
8. Mercer N (2000) *Words & Minds*. Routledge, London
9. Mercer N (2002) Developing dialogues. In: Wells G, Claxton G (eds) *Learning for Life in the 21st Century*. Blackwell Publishers, Oxford, pp 141-153
10. Mercer N, Littleton K (2007) *Dialogue and the Development of Children's Thinking*. Routledge, London and New York
11. Mercer N, Wegerif R (1999) Is 'exploratory talk' productive talk? In: Light P, Littleton K (eds) *Learning with Computers: Analysing Productive Interaction*. Routledge, London and New York, pp 79-101
12. Wegerif R (2008) Reason and dialogue in education. In: van Oers B, Elbers E, Wardekker W, van der Veer R (eds) *The Transformation of Learning*. Cambridge University Press, Cambridge, pp 273-288
13. Ferguson R, Whitelock D, Littleton K (2010) Improvable objects and attached dialogue: new literacy practices employed by learners to build knowledge together in asynchronous settings. *Digital Culture and Education* 2 (1):116-136
14. Mercer N (2004) Sociocultural discourse analysis: analysing classroom talk as a social mode of thinking. *Journal of Applied Linguistics* 1 (2):137-168
15. Sándor Á, Kaplan A, Rondeau G Discourse and citation analysis with concept-matching. In: *International Symposium : Discourse and document (ISDD)*, Caen, France. Web publication: http://www.unicaen.fr/services/puc/article.php3?id_article=686, 2006.

