Unraveling Idea Development in Discourse Trajectories

Iassen Halatchliyski\textsuperscript{1}, Aileen Oebers\textsuperscript{1}, Martina Bientzle\textsuperscript{1}, Franziska Bokhorst\textsuperscript{1}, Jan van Aalst\textsuperscript{2}

\textsuperscript{1}Knowledge Media Research Center, Schleichstr. 6, 72076 Tübingen, Germany
\textsuperscript{2}Faculty of Education, University of Hong Kong, 323 R S Building, Pokfulam Road, Hong Kong, China

Abstract: With the present paper we want to shed light onto an issue that is central within the knowledge building theory but only little studied – the development of ideas in collaborative learning discourse. Starting from the construction of a network of explicit and implicit relations between ideas, we apply a scientometric method to tackle the temporality of collaborative processes based on the structure of successive ideas. The resulting discourse trajectories are shown to give a holistic and also a detailed view on how knowledge advances when their interpretation is combined with a qualitative analysis of the content of the ideas and their relations. The weighted relevance of relations between ideas enables the identification of sub-topics in the discourse, important ideas, and influence or uptake events.

Introduction

How is knowledge about the world created and advanced? Knowledge building is an approach from the learning sciences that builds on contemporary philosophical views and research on expertise (Bereiter, 2002; Scardamalia & Bereiter, 2006) to engage students in modern knowledge work, including ability to collaborate, deal with novelty, and solve ill-structured problems. At the heart of knowledge building is a computer-mediated collaborative discourse that is oriented toward idea improvement. Following Popper’s (1972) theory of objective knowledge, knowledge-building theory considers ideas as “real” objects that can be critiqued, tested and modified, much like how real objects like bicycles undergo these processes. Hence, the process of idea development is fundamental for understanding knowledge building.

However, despite this acknowledged role, there is a dearth of analytical approaches for investigating the dynamic development of ideas in knowledge-building discourse. Therefore, the main goal of this paper is to present a methodological technique from scientometrics for studying what we call a discourse trajectory, i.e. the genuine process characteristics of a discourse based on idea development over time.

Related Research

Most studies of knowledge building have followed a content analysis approach to studying a discourse process (Gunawardena, Lowe, & Anderson, 1997; Henri, 1992), where qualitative data is segmented into idea units and these are coded for their cognitive, social and other aspects. The frequency of the assigned codes is then statistically regarded for comparisons of different students, discussion groups or phases. Students’ contributions can also be categorized as different discourse activities, on the basis of which a sequential analysis (Jeong, 2005) of their temporal ordering can reveal patterns and facilitate deeper insight into the collaborative discourse as a process. However, as is becoming increasingly clear in CSCL research, the coding and counting technique neglects some important discourse qualities as it takes statements out of their context and generally addresses actions of individuals instead of the group as a whole (Strijbos & Stahl, 2007). Stahl, Koschmann and Suthers (2006) defined collaborative learning as an interactive process of shared meaning-making in a group and pleaded for exploratory and interpretative conversation analysis of case study narratives.

Henri (1992) addressed interactivity in a discourse process distinguishing independent from implicit interactive and explicit interactive contributions depending on if and how they refer to other contributions in the discourse. Later, Gunawardena et al. (1997) extended this view noting that in a knowledge constructing discourse all contributions can be linked to one or more other contributions and to the discussed topic. This marks a change from understanding interactivity as reference between contributions to treating it as diffusion of knowledge and other more general forms of uptake or influence. Suthers, Dwyer, Medina and Vatrapu (2010) introduced and employed contingency graphs for uptake analysis with the goal to identify and map also very subtle kinds of contribution uptakes that often remain unnoted. As noted by Lipponen, Rahikainen, Lallimo and Hakkarainen (2003) a discussion thread structure conceals the majority of semantic and conceptual relations between contributions.

Another approach to studying interactivity in discourse processes presents the social network analysis (SNA), which is a well established methodology in the CSCL research field (e.g. de Laat, Lally, Lipponen, & Simons, 2007; Refray, & Chanier, 2002). It has been used for studying relations between persons embedded in a network determining cohesiveness of the learning group and students’ relative positions. Following the reasoning of Stahl (2006) on intersubjective meaning-making in CSCL in terms of networks of references the discourse process can also be approached from the perspective of a network of collaboratively created artifacts.
There are accordingly a few notable examples of studies of automatically detected semantic networks of related contributions (Sha, Teplov, & van Aalst, 2010). Because network analysis is based on indicators of relations that lack deeper meaning in the discourse, it should be applied in combination with other in-depth and content related methods.

In sum, an appropriate methodology for studying discourse as a process has to be multi-faceted and address the temporal dimension of development; the interactivity between contributions and between participants; the content of the discourse (Arvaja et al., 2007).

The main goal of our study is to provide a primer for a novel methodology that tackles the trajectory of a discourse process analyzing the collaborative development of ideas over time. The approach presented here is rather simple and grounded on a network analysis of interrelated ideas. Based on a set of identified relations a main path analysis algorithm (Hummon & Doreian, 1993) assigns different weights to the relations according to their position in the network. This weighted network can be interpreted in order to identify the most influential ideas, idea paths, i.e. successions of related contributions, discourse trajectory, i.e. the overall structure of the paths over time, etc. The procedure of main path analysis stems from the scientometric research tradition that deals with citations between scholarly publications (e.g. Carley, Hummon, & Harty, 1993). In the present application of the analysis again relations between authored content are of interest. To the best of our knowledge, the present work is a pioneering attempt of applying scientometric methodology in the field of CSCL, and knowledge building in particular.

Main path analysis calculates a weight for each relation in a citation network according to the number of times the relation is used while tracing all chronological connections between all possible pairs of contributions (De Nooy, Mrvar, & Batagelj, 2005). So, a relation on idea paths is more important, the more often it serves as a link between preceding and subsequent ideas. Contributions that establishing such important relations often synthesize old ideas, add new knowledge and represent a basis for developing new ideas. A main path in the discourse trajectory is produced by the links with the highest weights.

De Nooy et al. (2005) reason that networks of scientific citations represent systems of knowledge flows. This fits our goal to analyze development of ideas in a knowledge building discourse. The resulting discourse trajectory is comparable to the development of a scientific field and can be analyzed for instances of integration, fragmentation, specialization or paradigmatic changes (De Nooy et al., 2005). Moreover, we see very promising possibilities for combining it with in-depth and content analysis methods.

We reanalyzed Knowledge Forum data from a study of two Grade 10 classes, one regular and the other honors, who investigated aspects of environmental issues; see Niu & van Aalst (2009) for details. The original study showed that both high and low-achieving students can sustain a knowledge building discourse.

The students worked on the Knowledge Forum software in groups of approximately eight over a period of three weeks. For the present study we compared the discourses of one group from each class. Both groups separately discussed ideas for handling the problems of the nuclear accident at Chernobyl. The honors group wrote 60 notes and the regular group 16 notes on this issue. We analyzed the log data and the content of the computer notes; the log files contain writing (saving a note) and reading (opening a note) events, which were recorded with a timestamp while using Knowledge Forum.

As we were interested in collaborative learning at the level of idea development in discourse it was important to take all ideas and all paths of influence over time into account. We adopted Lipponen’s (2000, p.15) definition that an “idea is a set of propositions that formed a coherent unit of meaning” and determined that a single note represents a suitable unit of analysis for our data (cf. Rourke, Anderson, Garrison, & Archer, 2001). Even though a note may present new information taken from an external source, almost always it is also connected to some previously stated ideas, as put forward by Gunawardena et al. (1997). Direct links between notes in a thread capture only a small part of all existing relations between notes, however. Thus, in order to capture relations between ideas, we made use of both explicit links and implicit connections by analyzing data from two different sources, the log files of the software as well as the content of the contributed notes.

Based on the timestamps two independent raters considered all the notes each student had read before writing a new note. The raters then compared the content of each note the student had read and the content of her own new note. Any disagreements between the raters about the evidence of connection between notes were discussed and solved unanimously. Implicit relations in the discussions of the Chernobyl issue were found for example between notes that addressed the same point—the fault of the designer of the reactor vs. that of the
technical operator; the radiation outburst; the increase in the incidence of cancer; the old covering of the remains of the reactor; financial support for the Ukraine; and many more other points. The complete qualitative analysis of the content relations between the notes is still in progress. The discourse trajectory results reported here are only based on the detection of connections.

We then applied the main path analysis procedure (Hummon & Doreian, 1999; Carley, Hummon, & Harty, 1993) to the resulting networks of both group discourses and obtained weights for each link based on its relative importance for all paths of idea development. This was done with a standard procedure for main path analysis in the Pajek software (Batagelj & Mrvar, 1999) for network analysis.

Results
Figures 1 and 2 depict the results of the main path analysis of both group discourses. The vertically layered view illustrates idea development during the different days of activity in class and at home. The numbered points represent different notes. Arrows represent some of the relatively important relations between notes with thicker arrows denoting more important relations, i.e. higher weights calculated with the main path analysis. The arrows are directed from an older to a newer note. It is important to mind that the isolated notes positioned on the left and on the right of the figures were also found by the raters to be related to other notes, but these relations received very low weights in the main path analysis. In order to identify the main idea paths in the discourse trajectories more clearly, Figure 1 and 2 show only relations over the arbitrary threshold weight of 0.05 for normalized weights between 0 and 1.

Both figures display single connected main idea paths, because there were no disparate discussions around different topics in the groups. However, both figures show different discourse trajectories at a first glance that remain to be characterized with the help of the contributed content.

Honors Group (Figure 1) The notes from the first day on (top of the figure) generate concurrent idea paths interrelating and stimulating one another; addressing background information on causes and on effects of the accident. In April they the development of these ideas seems to get focused into one path dealing with technological issues based on the causes of the accident. After April the idea paths in the honors group separate resulting in four largely independent lines of inquiry that are pursued until the last days of the course. The thickest path deals with futuristic solutions of neutralizing radiation and inspires the largest amount of participation. The remaining three paths discuss the covering of the destroyed reactor, the use of nuclear energy in general and the politics behind the accident.

Figure 1. Discourse trajectory of the honors group: notes and main idea relations
Egu Group (Figure) All the initial idea paths merge into a single note, 316, on April 1st that brings up the need for solutions after the basic facts are known. Further tracking of the discourse trajectory suggests that this pattern repeats within the regular group. Concurrent idea developments are short-lived and continuously meld into a single note on the same or the next day. In April 1st note proposes a solution for the polluted soil around the old reactor, i.e. digging it and disposing of it in outer space. Then, again, this central idea gives rise to various related idea paths concerning difficulties of the solution – most importantly – the money problem. Additional solutions and refinements emerge on April 5th and are then put to a vote by note 6.3. Note 3 bundles the focus to the solution of building a new covering for the old reactor. The game of proposing solutions and recognizing difficulties continues yielding tightly interrelated idea paths.

Figure 2. Discourse trajectory of the regular group: notes and main idea relations

Conclusion
The short comments on the results illustrate some initial ways of interpreting discourse trajectories obtained through main path analysis. With the present paper we pursued the goal to open up a field of possibilities for studying collaborative learning processes as we introduced a new method to the field. We showed that it handles the temporal perspective of idea development very well providing an objective measure of the relevance of ideas and their relations. These can then be examined more closely with regard to their contents in a mixed methods approach (see also Carley et al., 1993). The obtained discourse trajectory also provides a holistic view on the collaborative process. Although both analyzed groups were successful in knowledge building, they showed very different styles of collaboration regarding the convergence-divergence polarity (Halatchliyski, Kimmerle, & Cress, 2011). The regular group produced a large number of notes, and their main ideas were very tightly interrelated; convergence was maintained over the whole discourse trajectory. The honors group achieved a self-organized “division of labor” by building up a common understanding of the problems and then following divergent solution paths. It remains to be shown to what extend this is due to individual differences.

Our next goal is to complete the qualitative analysis of the data set in order to determine what kind of contributions and relations receive higher or lower weights, i.e. are more or less important in the discourse.

References
This study resulted from an exploration the authors began during the Summer School “Making Sense of Social Media”, co-sponsored by the Knowledge Media Research Center and Tuebingen ScienceCampus, August 1-4, 2011. The data were originally collected as part of a new Economy Collaborative Research grant from the Social Sciences and Humanities Research Council of Canada to Marlene Scardamalia (OIS University of Toronto). The authors thank Hui liu (Simon Fraser University) for permission to reanalyze the database.