Computer Supported Content Analysis

—— Challenges, research and developments

Ronghuai Huang & Yanyan Li
Beijing Normal University
Email: huangrh@bnu.edu.cn
Http://ksei.bnu.edu.cn
Why CSCL?

- Three issues about cooperation
  - defined as “acting together, in a coordinated way at work, or in social relationships, in the pursuit of shared goals…”
  - Is seen as central to our everyday lives
  - Cooperative learning is process driven
- A human group is a collection of individuals, who have interdependent relations, and who perceive themselves as a group that is recognised by non members
- People working cooperatively in CSCL environments do work in groups in complex ways
Why CSCL?

• What are the outcomes of cooperative learning?
  Cooperative methods lead to higher achievement than competitive or individualistic ones.
  Cooperative learning increase the positive affect of classrooms and students working cooperatively become more cooperative; they learn pro-social behaviours such as hot to get with others, how to listen and so on.
  Cooperative learning fosters knowledge about the learning process.
Possibilities vs. Wicked problems

- Possibilities: Collaborative technologies are shown to enhance student motivation, self-reflection, working with complex problems, and promote collaboration between learners.

- 'wicked problems': a "problem that can be characterized as an evolving set of interlocking issues and constraints in a constantly changing context".

- Basic problems of technical infrastructure and shortage of IT-trained staff.
School culture challenges of CSCL

- issues concerning the compatibility of CSCL with the curriculum and the organizational structure of the school (e.g. Cullen)
- many learners seem to have great difficulties in participating in collaborative inquiry activities if these are not highly structured and if they are not given clear instructions (e.g. Blake & Rapanotti; Ploetzner et al.).
- learners often do not reach a higher level of discussion and knowledge building (e.g. Lipponen et al.; Muukkonen et al.; Mäkitalo et al.).
- many learners seem to operate under the assumption that a knowledge building process in school environment is a kind of a "question-answer-game" (see Kynigos, Dimaraki & Trouki).
- In addition, teachers were shown to have difficulties in guiding a collaborative inquiry process (e.g. Rahikainen et al.).
Pedagogical challenges of CSCL

- We do not yet seem to fully understand how technology should be employed in order to best support collaborative learning and higher-level knowledge building in different educational settings. The challenges include:
  - a need for deeper knowledge about the kinds of activities that should accompany CSCL.
  - seek for an understanding of the best combination of CSCL and traditional, more individualistic instructional approaches (e.g. Muukkonen et al.).
  - unequal participation of the learners in computer supported collaborative learning. Some studies pointed out a tendencial exclusion of weaker or less motivated learners from computer-mediated discussions (e.g. Cullen, Lipponen et al.; Rahikainen et al.; Tapola et al.).
  - better understanding the kind of pedagogical support needed during computer supported collaborative learning (e.g. Salovaara & Järvelä; Tholander).
Solutions?

- one of the central challenges of research on CSCL will be the developing of pedagogical models and methodological approaches.
- In order to answer both the cultural and pedagogical challenges, it seems that we also need to explore further the nature of computer supported communication and inquiry itself.
- there has been a change in the research on CSCL to more detailed research on the characteristics of discourse and argumentation.
- Accompanying this process, there is also a need to develop new ways of assessing the learning outcomes in computer supported collaborative learning, because the traditional assessment methods are not necessarily able to show the benefits gained through this kind of learning (e.g. Karlgren).
## CSCL: space and time

<table>
<thead>
<tr>
<th>SPACE</th>
<th>TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-present</td>
<td>Synchronous</td>
</tr>
<tr>
<td>Distant</td>
<td>Asynchronous</td>
</tr>
</tbody>
</table>

- **Co-present** (Synchronous): People are present in the same location and share the activity in real-time.

- **Distant** (Asynchronous): People are not present in the same location and engage in the activity at different times or asynchronously.

- **Synchronous**: Activities occur simultaneously and require participants to be present at the same time.

- **Asynchronous**: Activities can occur at different times or at different locations, allowing for flexibility in participation.

[Knowledge Science & Engineering Institute, Beijing Normal University, http://ksei.bnu.edu.cn]
“Constructive Interaction” paradigm

Cognitive dimension

$\Delta I$: Problem representation

$\Delta K$: task, social relation

Social dimension

$\Delta I$: appropriation of tools

$\Delta K$: intercomprehension, coordination

$S$: situations

$I$: interaction

$K$: knowledge
A New Approach

- It’s still unclearly about how the group member implement Collaborative Knowledge Building in interaction. (Koschmann)
- We should study and understand categories of interactions and the mechanisms of negotiation to a much greater depth rather than only study collaboration in general. (Dillenburg & Baker)

Effect of Collaboration
=Post-test
— Pre-test

Process of Collaborative Knowledge Building
A New Approach – Content Analysis?

Collaborative Knowledge Building (CKB)

- How to collaborate?
- Effect of collaboration

- Analyzing CL’s characteristics and status (Computation & Modeling)
- Building effective supporting environment

Knowledge Science & Engineering Institute, Beijing Normal University, http://ksei.bnu.edu.cn
Materials in content analysis

- Coding the **content of documents** (like behavior coding)
  - any technique for making inferences by identifying special characteristics of messages *(written or oral)*
    - **artifacts** of social communications
  - information is condensed *(classified)* and made systematically comparable by applying a **coding scheme**
- any kind of **written document**
  - field notes from participant observation, letters, novels, **transcripts** of recorded communications (such as T.V shows, interviews, etc.)
The steps in content analysis

1. fully describe the **phenomenon** to be studies (e.g. portrayal of the elderly in the media)
2. select the **media** that will be used for data
3. derive coding categories
   - **choose categories**, e.g. status of character, physical attractiveness, context, etc.
   - **count** presence or absence of a category
   - place each piece into one of many categories (**forced choice**)
4. decide on a **sampling strategy** --you can’t count it all
5. **train the coders/raters** (**reliability** is important)
6. **analyze the data** (%’s, compare means and variances?)
Research Perspective

- CL (Collaborative Learning) & CKB (Collaborative Knowledge Building)
  - Interaction in e-Learning Environment is communication among group members mediated by computer and verbal information. Although its form is various, its essential function is Collaborative Knowledge Building, whose essential feature is Collaboration.
  - CKB, which is interpreted from genetic epistemology, is the activity that the collaborative community builds the shared understanding and forms the inter-dependent relationship by the interaction among members under the same learning and cultural background.
Research Aim

- To find features of CKB and to analyze status of CL
- Explore the approach to extract the above feature from interaction corpus. To design and develop the tools to implement auto analysis or semi-auto analysis of interaction.
- To improve CKB based on the result of interaction analysis. To design and develop CKB Supported Tools.
## Analysis Framework: Henri Model

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Definition</th>
<th>Examples of Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participative</td>
<td>Compilation of the number of messages or statements transmitted by one person or group</td>
<td>Number of messages&lt;br&gt;Number of statements</td>
</tr>
<tr>
<td>Social</td>
<td>Statement or part of statement not related to formal content of subject matter</td>
<td>Self-introduction&lt;br&gt;Verbal support&lt;br&gt;'I'm feeling great.......!'</td>
</tr>
<tr>
<td>Interactive</td>
<td>Chain of connected messages</td>
<td>'In response to Celine.....'&lt;br&gt;'As we said earlier......'</td>
</tr>
<tr>
<td>Cognitive</td>
<td>Statements exhibiting knowledge and skills relating to learning processes</td>
<td>Asking questions&lt;br&gt;Making inferences&lt;br&gt;Formulating hypotheses</td>
</tr>
<tr>
<td>Meta-cognitive</td>
<td>Statements related to general knowledge and skills and showing awareness, self-control, and self-regulation of learning</td>
<td>Commenting on own manner of accomplishing a task&lt;br&gt;Being aware of the emotional context of task completion</td>
</tr>
<tr>
<td>Analytical Category (after Henri, 1992)</td>
<td>Female Student</td>
<td>Male Student</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>----------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Social</td>
<td>26%</td>
<td>35%</td>
</tr>
<tr>
<td>Interactive</td>
<td>37%</td>
<td>26%</td>
</tr>
<tr>
<td>Cognitive Skills (surface)</td>
<td>11%</td>
<td>0%</td>
</tr>
<tr>
<td>Cognitive Skills (deep)</td>
<td>26%</td>
<td>39%</td>
</tr>
<tr>
<td>Meta-cognitive Knowledge</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Meta-cognitive Skills</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Phase I</td>
<td>Sharing/Comparing of Information</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>---</td>
</tr>
<tr>
<td>A</td>
<td>Statements of observation/opinion</td>
<td>6</td>
</tr>
<tr>
<td>B</td>
<td>Agreement from one or more participants</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>Corroborating examples from one or more participants</td>
<td>6</td>
</tr>
<tr>
<td>D</td>
<td>Clarification by asking or answering questions</td>
<td>9</td>
</tr>
<tr>
<td>E</td>
<td>Definition, description or identification of a problem</td>
<td>2</td>
</tr>
<tr>
<td>Phase II</td>
<td>Discovery and Exploration of Dissonance or Inconsistency</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Identifying and stating areas of disagreement</td>
<td>6</td>
</tr>
<tr>
<td>B</td>
<td>Asking and answering questions to clarify sources of disagreement</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>Restating a position and supporting it with evidence</td>
<td>1</td>
</tr>
<tr>
<td>Phase III</td>
<td>Negotiation of Meaning and Co-Construction of Knowledge</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Negotiation or clarification of the meaning of terms</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>negotiation of the relative weight to be assigned to types of argument</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Identification of areas of agreement or overlap among conflicting concepts</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Proposal and negotiation of new statements embodying compromise/co-construction</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Proposal of integrating or accommodating ideas</td>
<td></td>
</tr>
<tr>
<td>Phase IV</td>
<td>Testing and Modification of Proposed Synthesis/Co-construction</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Testing synthesis against established/received shared ideas of participants</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Testing against existing cognitive schema</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Testing against personal experience</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Testing against formally collected data</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Testing against contradictory testimony in the literature</td>
<td></td>
</tr>
<tr>
<td>Phase V</td>
<td>Agreement/application of New Co-constructions</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Summarisation of agreements</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Application of new knowledge</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Metacognitive statements indicating changes of understanding among participants</td>
<td></td>
</tr>
</tbody>
</table>
Research framework on CKB Process

- **Topic Space**: Topic (knowledge) set in process of CKB.
- **Social Network**: Relationships among group members.
- **Process Pattern**: The relatively steady path and organizing form of CKB.

**TSP Model**

- **T**opic Space: Topic (knowledge) set in process of CKB.
- **S**ocial Network: Relationships among group members.
- **P**rocess Pattern: The relatively steady path and organizing form of CKB.
Topic Space:

Social Relationship:

Process Pattern:
Research Approach
Vinca Introduction

- VINCA stands for *Visual INtelligent Content Analyzer*, which is the content analysis tool jointly developed by CITE, HKU and KSEI, BNU.
Vinca ‘s Features

- Learnable Semi-automatic Coding Support
- Analyze text in Chinese
- Utilize Computational Language & Text Mining technologies
- Support assessing for CKB
Vinca’s Functions

- Data preparation
- Annotation Aids
- Text Analysis
- Data Export for SNA
Data preparation

- Data preparation to convert Knowledge Forum discourse in html to database format
  - From Version 3.4
  - From Version 4.5
Annotation Aids

- **Edit Coding Scheme**
  - New, Modify, Delete
    - Associate feature keywords to specific codes
- **Annotation**
  - Automatic discover the code hint, highlight it and attach possible codes with confidence probability.
  - Support segment & merge
  - During the process of coding, users are allowed to select the hint to mark the final coding.
- **View Coding Result**
Select source with multiple modes
Associate feature keywords with specific codes

Edit code scheme
Tree-structure coding scheme

Coded notes are marked with blue

Auto highlighting the hint

Allow users to segment one note into multiple segment and vise versa

Knowledge Science & Engineering Institute, Beijing Normal University, http://ksei.bnu.edu.cn
View the coding result
Text Analysis

- Keywords retrieval & frequency counting
- Concordance
- Domain ontology-based category analysis
- Text Clustering
- Support assessing for CKB
<table>
<thead>
<tr>
<th>ID</th>
<th>Word</th>
<th>Count</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>伟</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>胜</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>教师</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>邮箱</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>模式</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>理论</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>仿</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>促进</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>不</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>用来</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>命</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

**Group Performance**

- **Keywords Count**: 396
- **Ontology Count**: 44

**Domain ontology**

**Member Contribution**

<table>
<thead>
<tr>
<th>Member</th>
<th>Contribution</th>
<th>Relevance</th>
<th>Novelty</th>
<th>Extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>李佳</td>
<td>0.375814</td>
<td>0.4299406</td>
<td>0.2890173</td>
<td>0.390422</td>
</tr>
<tr>
<td>姚浩</td>
<td>0.3705473</td>
<td>0.4601308</td>
<td>0.2745085</td>
<td>0.387052</td>
</tr>
<tr>
<td>罗惠</td>
<td>0.3396292</td>
<td>0.4269859</td>
<td>0.2613709</td>
<td>0.3192296</td>
</tr>
<tr>
<td>刘梅</td>
<td>0.3301135</td>
<td>0.4500396</td>
<td>0.1670167</td>
<td>0.358753</td>
</tr>
</tbody>
</table>

**Topic similarity between members**
Data Export for SNA

- Export KF Data
- Export Relation Matrix
- Export Coding Result
- Export Coding Matrix
- Export Coding Frequency
Thanks

Welcome Questions and Comments