



Computerized Discourse Analysis of TIMMS-R Lesson Transcripts

Bruce Lambert, Ph.D., Clement Yu, Ph.D.,
David Lewis, Ph.D., Fang Liu, & Rodica
Waivio

University of Illinois at Chicago
in association with LessonLab, Inc.



Overview

- TIMMS-R International Video Study
- Manual Content Analysis
- Computer-Assisted Content Analysis
 - Descriptive analyses
 - Comparative analyses
 - Clustering
 - Classification



TIMMS-R Video Study

- Project led by LessonLab, Inc. (www.lessonlab.com)
- Goal: to compare lesson form and content across cultures
- Replication, with 7 countries, of TIMMS study
- US, Hong Kong, Czech Republic, Australia, Switzerland, Japan, Netherlands
- Video tape of 100 math and 100 science lessons (eighth grade)



TIMMS-R Video Study

- Lessons transcribed and translated to English
- Lessons extensively coded via manual content analysis
- Detailed analysis of discourse form and structure done by computer-assisted means



Excerpt of Transcribed Lesson

M-AU-002-T T IP_1 00:36:58:04 Okay, it's a little bit different with the diameter, because we have to first of all work out what is the radius.

M-AU-002-T T IP_1 00:37:04:15 Lyle, what's the radius in this case?

M-AU-002-T SN IP_1 00:37:06:00 I am the walrus.

M-AU-002-T SN IP_1 00:37:08:08 Seven? Eight? Eight?

M-AU-002-T T IP_1 00:37:10:01 Good. Okay, our formula is $P = 2\pi R$ squared. Shhh.

M-AU-002-T SN IP_1 00:37:18:12 Can't you just do it with the perim - with the perimeter?

M-AU-002-T T IP_1 00:37:22:01 No, you have to do it using just the radius. So the radius is eight.

M-AU-002-T T IP_1 00:37:32:20 Calculators here, please. Good work. Good. Good, Jerry.

M-AU-002-T T IP_1 00:37:44:07 Okay. Two examples.

M-AU-002-T SN IP_1 00:37:51:08 Is it always going to be - is it always squared?



Computerized Text Analyses

- Phase 1: Word counts, concordances, sorting, global search and replacement
- Phase 2: Text similarity computation and document clustering
- Phase 3: Text categorization



Phase 1: Word counts, concordances, sorting, and search and replacement.

- Tools: GNU Emacs, GNU text utilities (sed, comm, sort, etc.), Gawk, Lisp
- Challenges
 - Translation issues
 - Transcription issues
 - Inaudible turns
 - Segmentation of transcript into turns



Phase 2: Text similarity computation and document clustering

- Tools: ThemeMachine document clustering system, plus refinements
- Steps:
 - Transform documents into vectors of term weights
 - Compute interdocument similarity via vector correlation
 - Hierarchical agglomerative clustering



Clustering Results

- Using verbatim text, lessons appear to cluster into distinct country- or culture-specific clusters
- Using synonym replacement and similarity measures based on conceptual hierarchy, country-specific clustering appears less pronounced



Clustering Challenges

- How much pre-processing?
 - Stop words?
 - Stemming?
- Which clustering parameters/methods?
 - Binary or IDF term weights?
 - Complete linkage or group-average clustering
- How to display clusters?
- Interpretation?



Phase 3: Text categorization

- Tools: See5 classification program, custom-designed programs
- Steps:
 - Select target code
 - Identify/create textual features
 - Transform textual units into fixed, attribute-value vectors
 - Train and evaluate classifier



Categorization Challenges

- Which codes to automate?
- What textual features should be used for classification?
- How accurate must automated coding be before it replaces human coding?
- How to integrate human and machine coding?
- How to interpret classification models?



Discussion and Implications

- Computer content analysis will enable a level of detailed description not previously possible
- Text classification methods may be able to supplant manual coding in certain circumstances
- Clustering may help to answer questions about the existence and nature of “cultural models” of classroom interaction



Acknowledgments

- Jim Stigler, Ron Gallimore, Sam Mansukhani, Ken Mendoza, Helen Garnier, Genevieve Patthey-Chavez, Jennifer Jacobs, LessonLab, Inc.