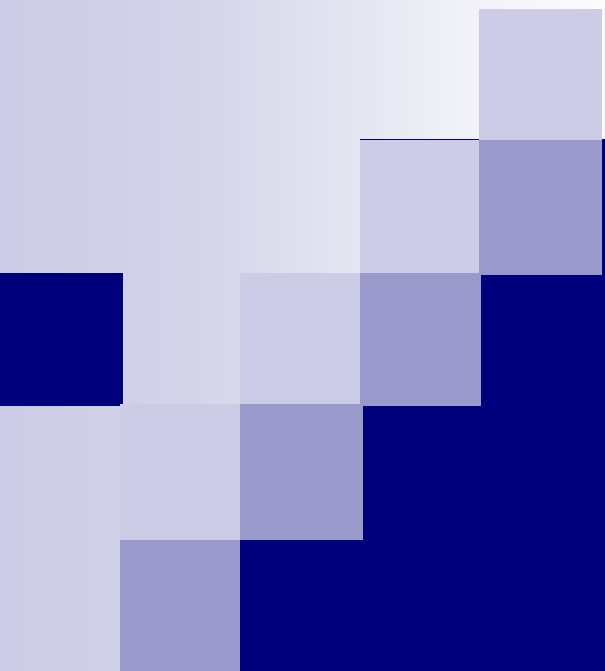




Discourse Analytics and Educational Data Mining

Carolyn Penstein Rosé

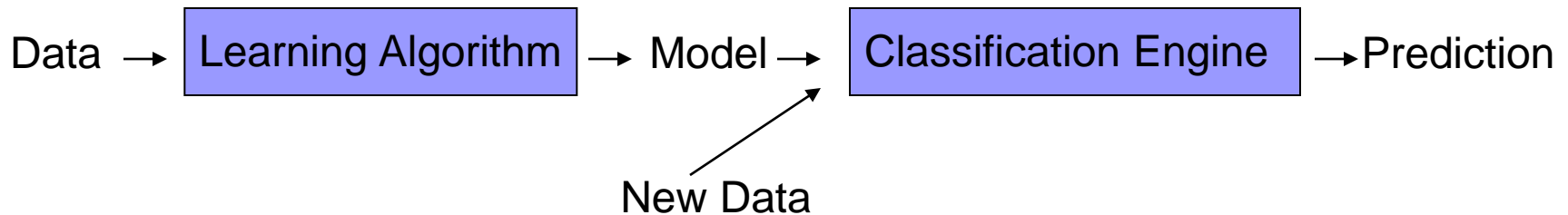
*Language Technologies Institute/
Human-Computer Interaction
Institute*



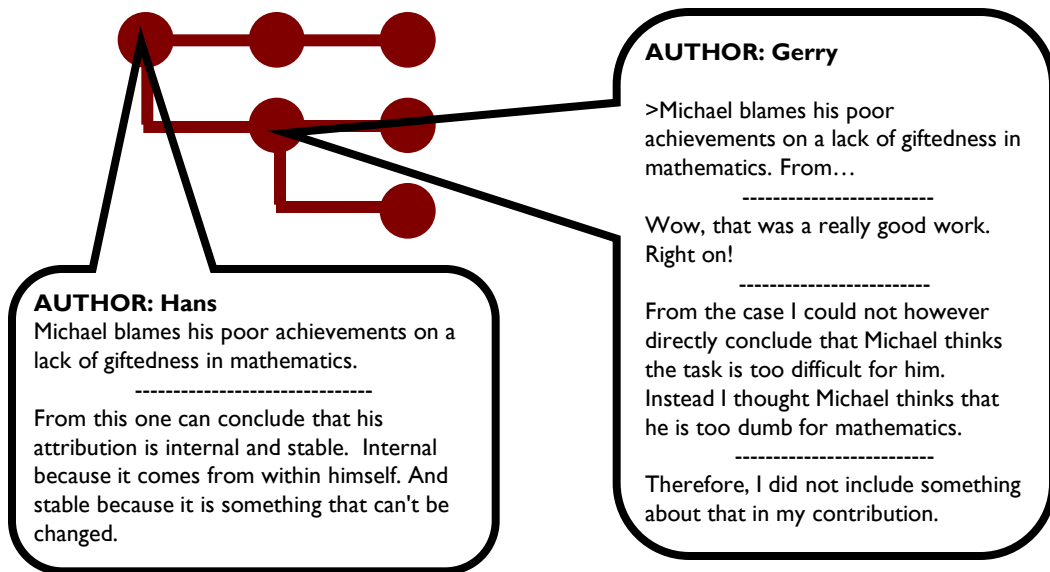
Overview of TagHelper Article Results

What is machine learning?

- Automatically or *semi-automatically*
 - Inducing rules from data
 - Making predictions



Identifying Transactivity in Threaded Discussions

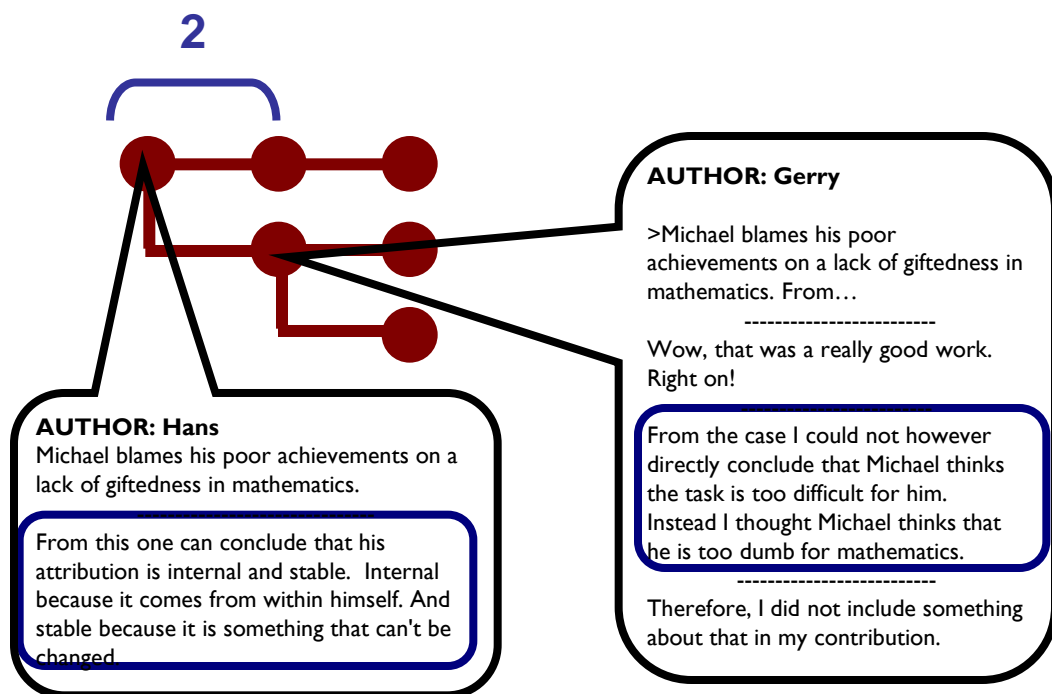


- Social modes of co-construction (Weinberger & Fischer, 2006)
 - To what degree or in what ways learners refer to the contributions of their learning partners
- TagHelper tools achieves reliability of .69 Kappa (Rosé et al., 2008)

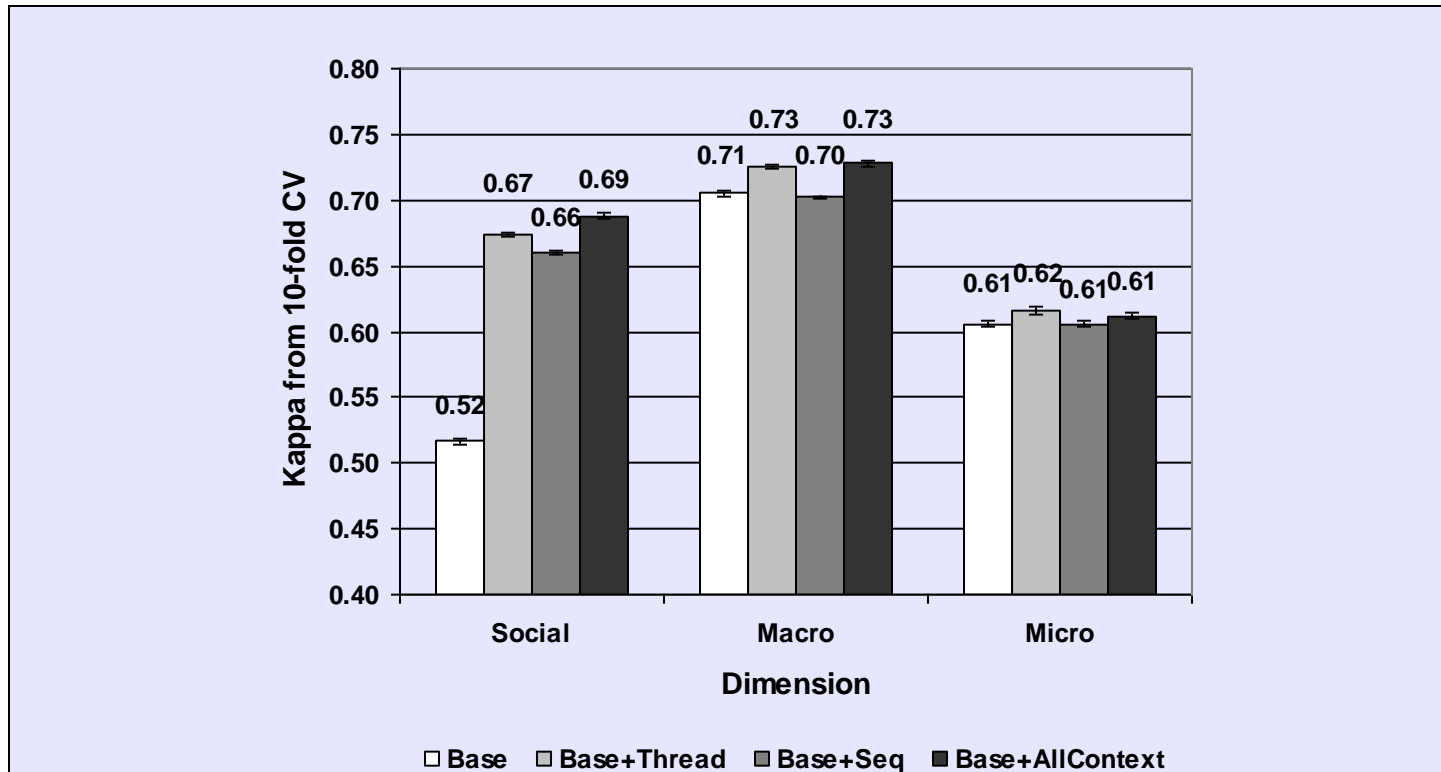
Thread Structure Features

- Thread structure features

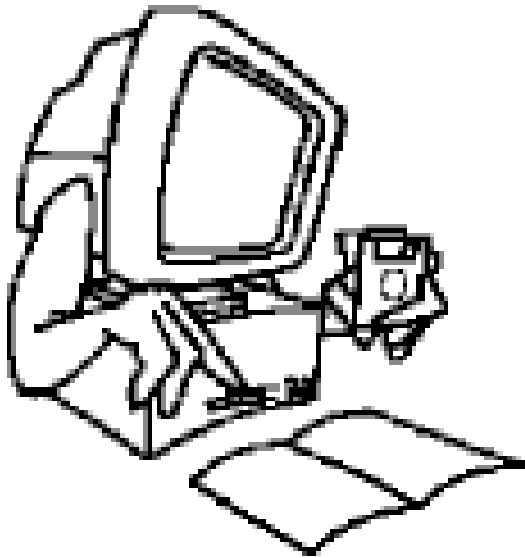
- **depth** (numeric): the depth in the thread where a message appears
- **parent_child_similarity** (numeric): semantic similarity (cosine similarity) between the current message segment to all its parent message segments. The highest value is chosen



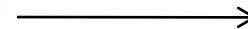
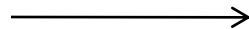
Evaluating Context-Based Features



Important caveat!!



- Machine learning isn't magic
- But it can be useful for identifying meaningful patterns in your data when used properly
- Proper use requires insight into your data





Discussion Questions

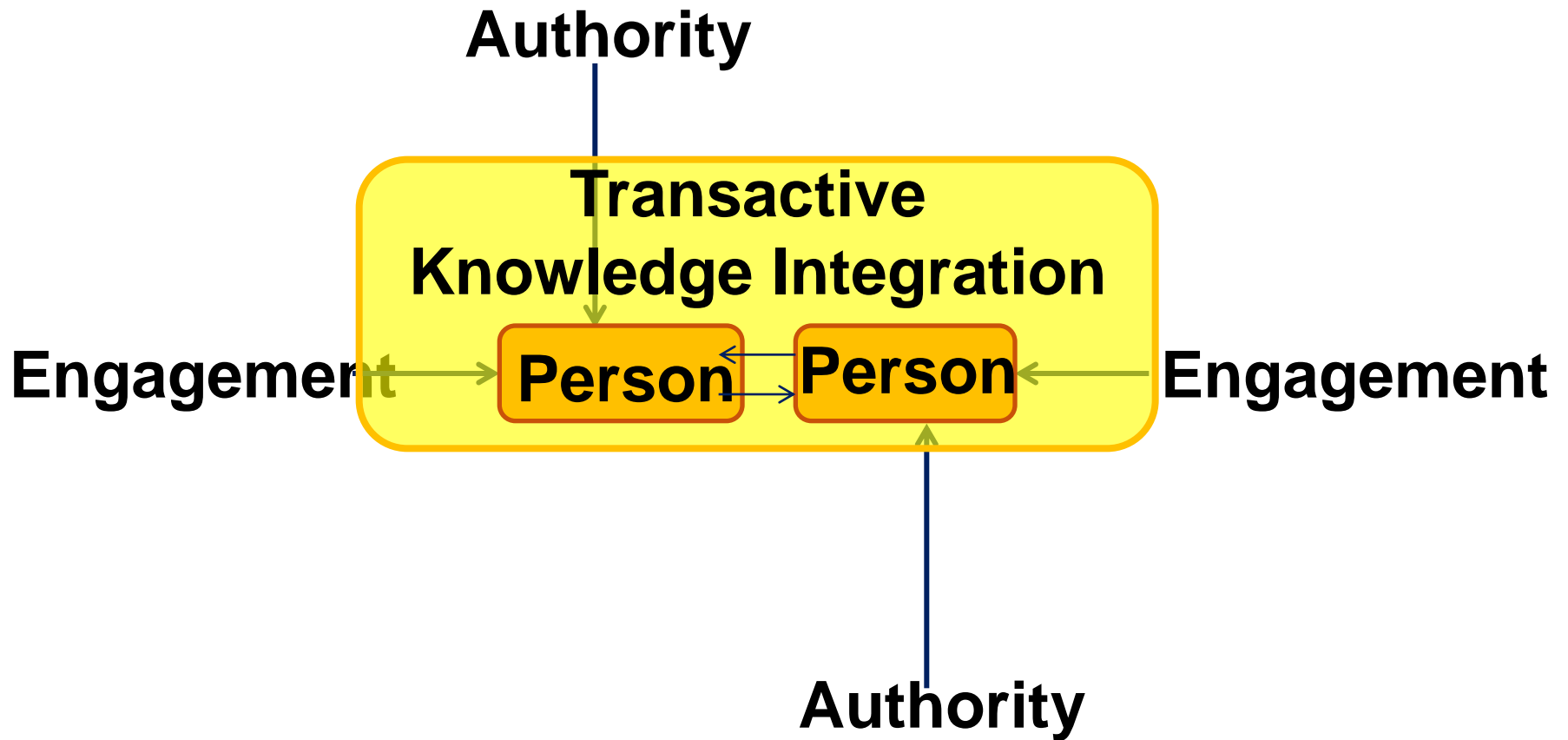
- Do you believe the results from this paper are strong enough to convince you to use automated analysis in your work?
 - If not, what would you need to see?
- How would you summarize the issues with respect to reliability and validity when using automated coding in comparison with hand coding?
- In what specific ways could you imagine using automated coding in your own research?



SouFLé (part 1)

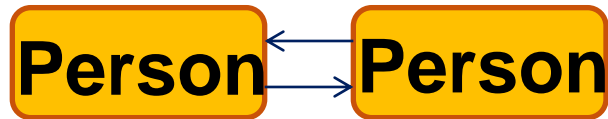
Transactivity

SouFLé Framework (Howley et al., 2013)



Souflé Framework

(Howley et al., in press)



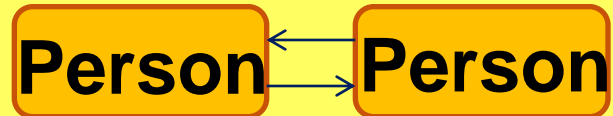
3 Dimensions:

- Transactivity
- Engagement
- Authoritativeness

Souflé Framework

(Howley et al., in press)

**Transactive
Knowledge Integration**



• Definition of Transactivity

- building on an idea expressed earlier in a conversation
- using a reasoning statement

I think the tube will get heavier because water is going in

That's true, but the important point is that water can flow in, but starch can't flow out.

Transactivity (Berkowitz & Gibbs, 1983)

■ Findings

- Moderating effect on learning (Joshi & Rosé, 2007; Russell, 2005; Kruger & Tomasello, 1986; Teasley, 1995)
- Moderating effect on knowledge sharing in working groups (Gweon et al., 2011)

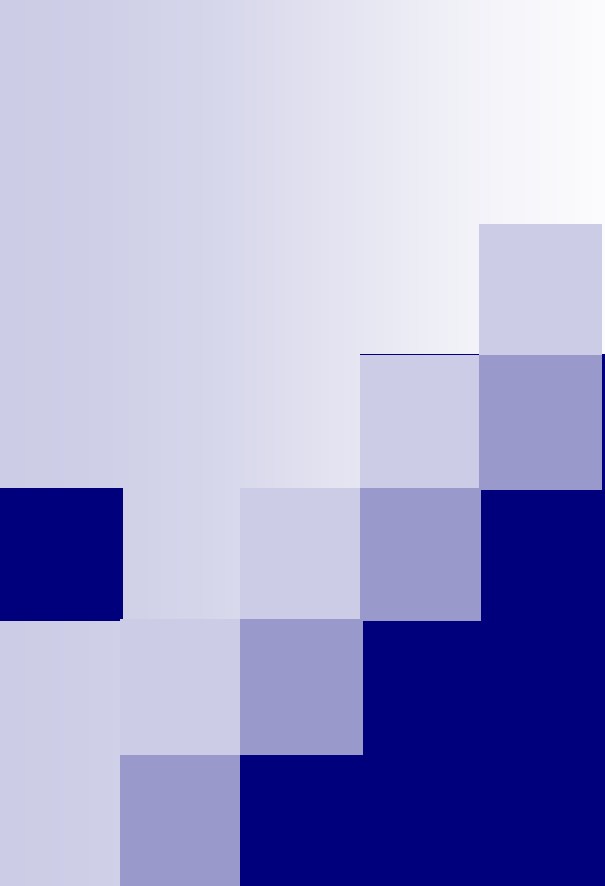
■ Computational Work

- Can be automatically detected in:
 - Threaded group discussions (Kappa .69) (Rosé et al., 2008)
 - Transcribed classroom discussions (Kappa .69) (Ai et al., 2010)
 - Speech from dyadic discussions (R = .37) (Gweon et al., 2012)
- Predictable from a measure of speech style accommodation computed by an unsupervised Dynamic Bayesian Network (Jain et al., 2012)

			Reasoning	Transactivity
AUTHOR	TEXT	TIME		
Doctor Bob	Hi - I'm Doctor Bob, your instructor for today's lab.	05.07.13		
Sa04	Well bob, I hope you don't mind that I'm actively hostile to robots.	05.07.36		
Doctor Bob	You're working together, but you've each got a different design goal - it's at the top of	05.07.39		
Doctor Bob	Take a moment to share your goal with your partner.	05.07.42		
sa08	most environmentally friendly	05.08.00		
Sa04	my goal is to maximize power. yours?	05.08.10		
Doctor Bob	To help you meet your design goals, we are going to discuss some of the parameters	05.08.43		
Doctor Bob	Lets start with Tmax.	05.08.46		
sa08	so basically, we compromise on greenness and power	05.10.53		
Doctor Bob	The heat rejected by the cycle increases by increasing Tmax.	05.11.13	X	
Doctor Bob	We see this in Graph 3 of the worksheet.	05.11.16		
Doctor Bob	BTW: Is it safe to keep increasing Tmax?	05.12.17		
sa08	no	05.12.33		
Sa04	nope. you'll blow out your turbine	05.12.33	X	
Doctor Bob	Good. Material properties constrain the maximum temperature we can use in a Cycle. For our cycle, Tmax cannot be more than 570C	05.12.39	X	X
Doctor Bob	Try using this understanding that our team now shares to come up with a potential value of Tmax (T @ S2) that will help you meet your objectives.	05.12.44		
sa08	We don't want it to be at 570 both for the material and [the environment]	05.12.48	X	X
sa08	well, for power and efficiency, we want a high tmax, but environmentally, we want a lower one. It has to be higher than 410 for steam quality	05.14.03	X	X
sa08	so somewhere between 410 and 570	05.14.26	X	X
sa08	what about right in the middle, what about 500?	05.14.40		
Sa04	seems reasonable	05.14.57		
sa08	We choose 500 degrees C	05.15.21		
Sa04	however, environmental friendliness can be increased by either increasing efficiency or by reducing waste heat, so maybe it's better to just max out our temperature.	05.15.53	X	

Discussion Questions

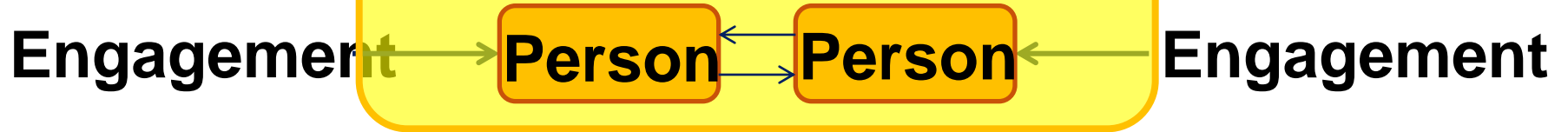
- How does the construct of transactivity relate to conversational or non-conversational constructs you have investigated in your own work?
- Based on your understanding of transactivity, which of the following would you expect to correlate with it and why:
 - Level of rapport within groups
 - Usage of dialect specific language features
 - Gesture and gaze
- How might non-linguistic and extra-linguistic features that correlate with transactivity be used in an automated conversation analysis approach?



SouFLé (part 2)

Engagement/Heteroglossia

Transactive Knowledge Integration



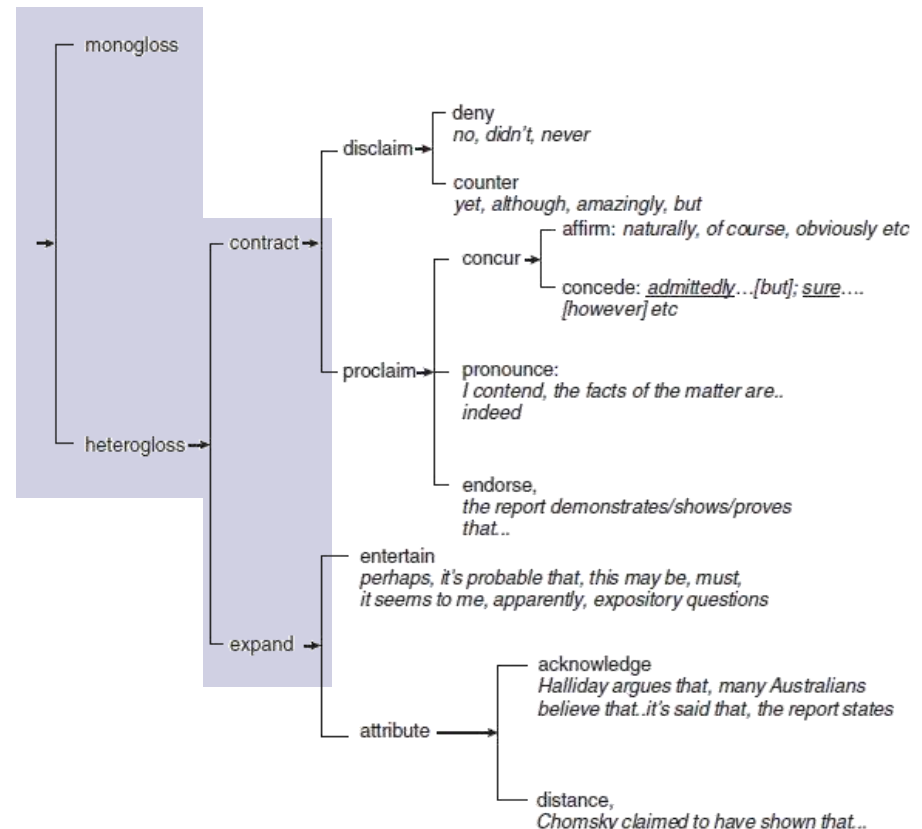
- System of Engagement
 - Showing openness to the existence of other perspectives
 - Less final / Invites more discussion

□ Example:

- **[M]** Nuclear is a good choice
- **[HE]** I consider nuclear to be a good choice
- **[HC]** There's no denying that nuclear is a superior choice
- **[NA]** Is nuclear a good choice?

Engagement

(Martin & White, 2005, p117)



Engagement (Martin & White, 2005)

■ Findings

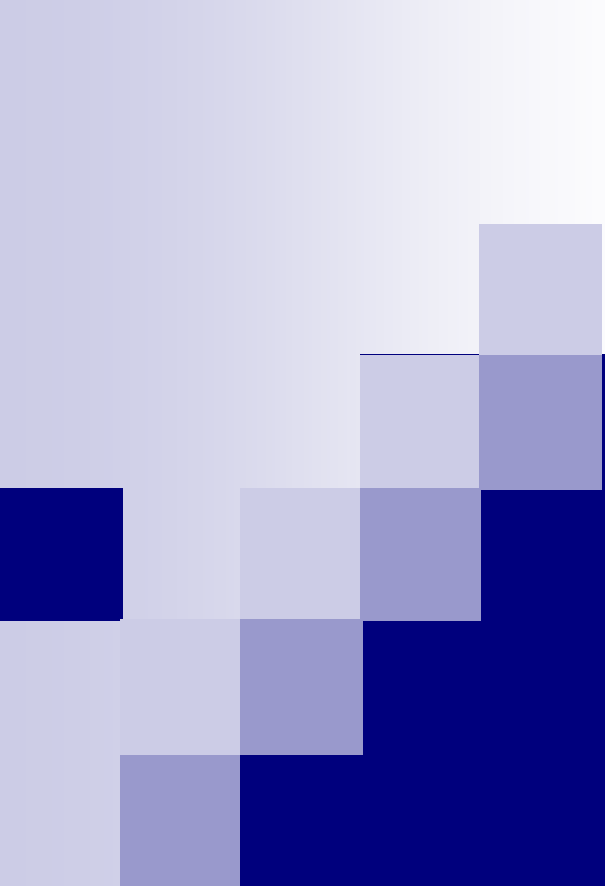
- *Correlational analysis*: Strong correlation between displayed openness of group members and articulation of reasoning ($R = .72$) (Dyke et al., 2013)
- *Intervention study*: Causal effect on propensity to articulate ideas in group chats (effect size .6 standard deviations) (Kumar et al., 2011)
 - Mediating effect of idea contribution on learning in scientific inquiry (Wang et al., 2011)

AUTHOR	TEXT	TIME	Engagement
Doctor Bob	Hi - I'm Doctor Bob, your instructor for today's lab.	05.07.13	[M]ono
Sa04	Well bob, I hope you don't mind that I'm actively hostile to robots.	05.07.36	[HE]xpand
Doctor Bob	You're working together, but you've each got a different design goal - it's at the top of	05.07.39	[M]ono
Doctor Bob	Take a moment to share your goal with your partner.	05.07.42	
sa08	most environmentally friendly	05.08.00	[M]ono
Sa04	my goal is to maximize power. yours?	05.08.10	[M]ono
Doctor Bob	To help you meet your design goals, we are going to discuss some of the parameters	05.08.43	[M]ono
Doctor Bob	Lets start with Tmax.	05.08.46	
sa08	so basically, we compromise on greenness and power	05.10.53	[M]ono
Doctor Bob	The heat rejected by the cycle increases by increasing Tmax.	05.11.13	[M]ono
Doctor Bob	We see this in Graph 3 of the worksheet.	05.11.16	[HE]xpand
Doctor Bob	BTW: Is it safe to keep increasing Tmax?	05.12.17	
sa08	no	05.12.33	[M]ono
Sa04	nope. you'll blow out your turbine	05.12.33	[M]ono
Doctor Bob	Good. Material properties constrain the maximum temperature we can use in a Cycle. For our cycle, Tmax cannot be more than 570C	05.12.39	[M]ono
Doctor Bob	Try using this understanding that our team now shares to come up with a potential value of Tmax (T @ S2) that will help you meet your objectives.	05.12.44	
sa08	We don't want it to be at 570 both for the material and [the environment]	05.12.48	[HE]xpand
sa08	well, for power and efficiency, we want a high tmax, but environmentally, we want a lower one. It has to be higher than 410 for steam quality	05.14.03	[HE]xpand
sa08	so somewhere between 410 and 570	05.14.26	[M]ono
sa08	what about right in the middle, what about 500?	05.14.40	[HE]xpand
Sa04	seems reasonable	05.14.57	[HE]xpand
sa08	We choose 500 degrees C	05.15.21	[HE]xpand
Sa04	however, environmental friendliness can be increased by either increasing efficiency or by reducing waste heat, so maybe it's better to just max out our temperature.	05.15.53	[HE]xpand



Discussion Questions

- How does the construct of Engagement relate to conversational or non-conversational constructs you have investigated in your own work?
- What might make Engagement easier or harder to recognize automatically than transactivity?
- How would you explain the correlation between Engagement and articulation of Reasoning in discussions?



SouFLé (part 3)

Authoritativeness

Authority

**Transactive
Knowledge Integration**

Person ↔ **Person**

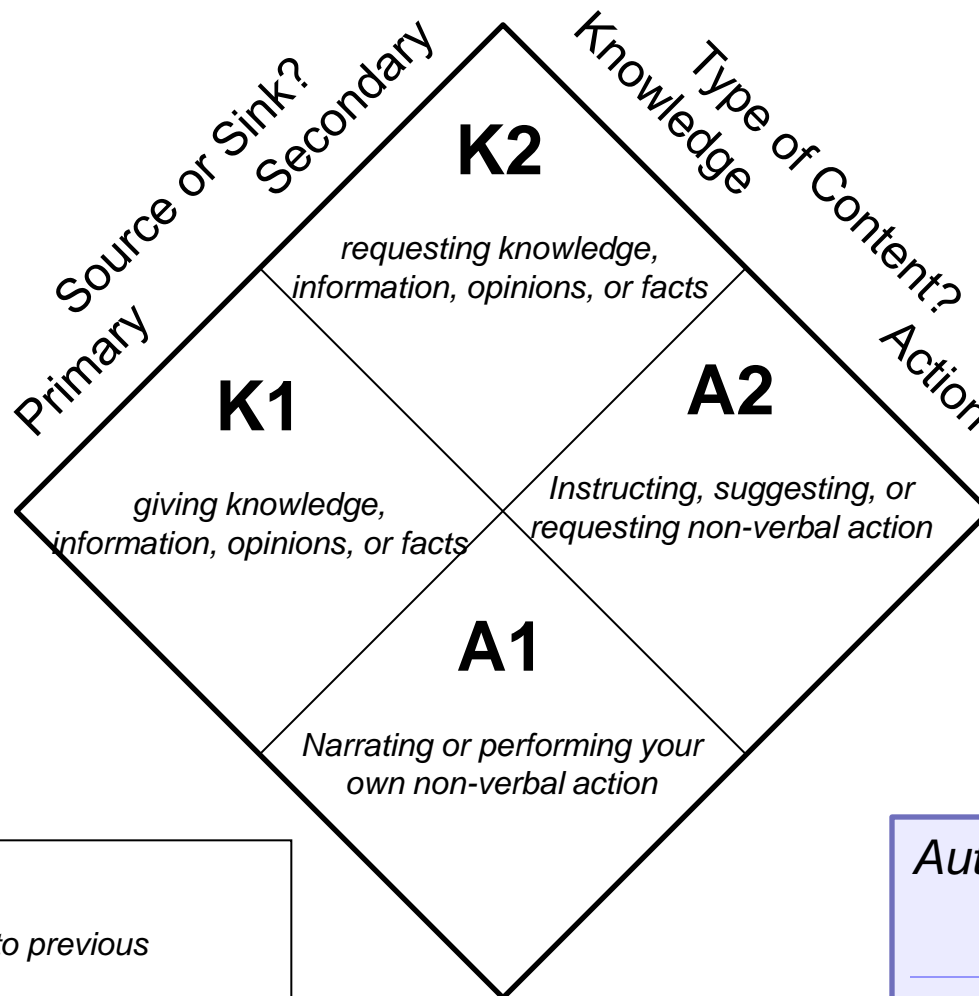
Engagement

Engagement

Authority

The Negotiation Framework

(Martin & Rose, 2003)



Additionally...

ch (*direct challenge to previous utterance*)

o (*all other moves, backchannels, etc.*)

Authoritativeness:

K1 + A2

K1 + K2 + A1 + A2

Authoritativeness (Martin & Rose, 2003)

■ Findings

- Authoritativeness measures display how students respond to aggressive behavior in groups (Howley et al., in press)
- Authoritativeness predicts learning ($R = .64$) and self-efficacy ($R = .35$) (Howley et al., 2011)
- Authoritativeness predicts trust in doctor-patient interactions (R values between .25 and .35) (Mayfield et al., under review)

■ Computational Work

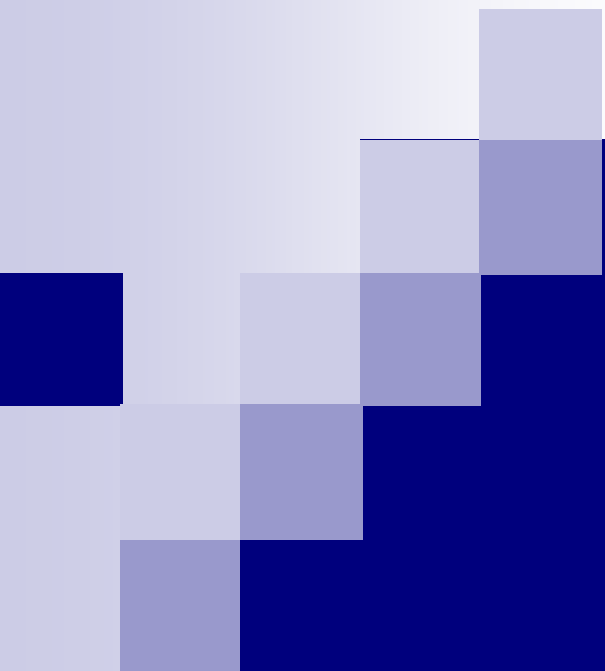
- Detectable in collaborative learning chat logs ($R = .86$)
- Detectable in transcribed dyadic discussions in a knowledge sharing task ($R = .95$) (Mayfield & Rosé, 2011)
- Detectable in transcribed doctor-patient interactions ($R = .96$) (Mayfield et al., under review)

			Authority
AUTHOR	TEXT	TIME	
Doctor Bob	Hi - I'm Doctor Bob, your instructor for today's lab.	05.07.13	K1
Sa04	Well bob, I hope you don't mind that I'm actively hostile to robots.	05.07.36	K1
Doctor Bob	You're working together, but you've each got a different design goal - it's at the top of	05.07.39	K1
Doctor Bob	Take a moment to share your goal with your partner.	05.07.42	A2
sa08	most environmentally friendly	05.08.00	K1
Sa04	my goal is to maximize power. yours?	05.08.10	K1/ K2
Doctor Bob	To help you meet your design goals, we are going to discuss some of the parameters	05.08.43	K1
Doctor Bob	Lets start with Tmax.	05.08.46	A1
sa08	so basically, we comprmise on greeness and power	05.10.53	A1
Doctor Bob	The heat rejected by the cycle increases by increasing Tmax.	05.11.13	K1
Doctor Bob	We see this in Graph 3 of the worksheet.	05.11.16	K1
Doctor Bob	BTW: Is it safe to keep increasing Tmax?	05.12.17	O
sa08	no	05.12.33	K2
Sa04	nope. you'll blow out your turbine	05.12.33	K2
Doctor Bob	Good. Material properties constrain the maximum temperature we can use in a Cycle. For our cycle, Tmax cannot be more than 570C	05.12.39	K1
Doctor Bob	Try using this understanding that our team now shares to come up with a potential value of Tmax (T @ S2) that will help you meet your objectives.	05.12.44	A2
sa08	We don't want it to be at 570 both for the material and [the environment]	05.12.48	K1
sa08	well, for power and efficiency, we want a high tmax, but environmentally, we want a lower one. It has to be higher than 410 for steam quality	05.14.03	K1
sa08	so somewhere between 410 and 570	05.14.26	K1
sa08	what about right in the middle, what about 500?	05.14.40	K1
Sa04	seems reasonable	05.14.57	o
sa08	We choose 500 degrees C	05.15.21	A1
Sa04	however, environmental friendliness can be increased by either increasing efficiency or by reducing waste heat, so maybe it's better to just max out our temperature.	05.15.53	K1



Discussion Questions

- How does the construct of Authoritativeness relate to conversational or non-conversational constructs you have investigated in your own work?
- How would you explain the connection between Authoritativeness and Self-efficacy? Would you be surprised if there was a correlation in learning contexts but not for doctors in doctor-patient interactions?
- What might explain the correlation between Authoritativeness and learning?



LightSIDE: A more powerful tool for Automated Process Analysis

Automated analysis – Quick Tour

The screenshot shows the LightSIDE Labs website. At the top, the browser address bar displays "lightsidelabs.com/research/". The main header features the "LightSIDE Labs" logo, where the 'i' in "LightSIDE" is replaced by a stylized yellow lightbulb icon. Below the logo is a dark navigation bar with links for "Home", "Machine Learning", "Assessment", "Tutorials", "Hiring", and "Contact Us", along with a search box.

The main content area is titled "Machine Learning" and features a large black button with the text "Download LightSIDE Now" and "Latest Update 5.10.2013". Below this, a paragraph states: "We're happy that you're interested in using LightSIDE as part of your research! Here's what the freely available and completely GPLv3 open source version of LightSIDE does for you:"

Underneath is a section titled "Easy, Fast Feature Extraction" which displays a screenshot of the LightSIDE software interface. The interface includes a "CSV Files" section with a file named "sentiment_sentences", a "Feature Extractor Plugins" section with options for "Basic Features", "Column Features", and "Previous Label Features", and a "Configure Basic Features" section with various checkboxes like "Ingrams", "Bigrams", "POS Systems", "Line Length", "Caseless Non-Whitespace", "Binary N-grams?", "Include Punctuation?", "Remove stopwords?", "Stem N-grams?", and "Differentiate text columns?". At the bottom, there are sections for "Feature Table" and "Evaluations to Display".

On the right side of the website, there is a search box and two sections: "Recent Posts" and "Recent Comments".

Recent Posts

- Survey: LightSIDE satisfaction
- Tutorial: Quick Start to Error Analysis
- Tutorial: Regular Expression and Stretchy Pattern Features
- Tutorial: Saving and Exporting Results
- Tutorial: Model Evaluation Settings

Recent Comments

- Rick Weinberg on Contact Us
- Rick Weinberg on Machine Learning
- Amarnath on Tutorial: Preparing My Data
- elijah on Contact Us
- Verb Washington on Contact Us

Archives

- April 2013
- February 2013

Categories

- Uncategorized

Meta

- Log In
- Entries RSS

CSV Files:

sentiment_sentences_cv.csv

DOCUMENT_LIST

Documents: sentiment_sentences_cv.

Class: class

Type: NOMINAL

Text Fields:

text

Differentiate Text Fields

Feature Extractor Plugins:

- Basic Features
- Character N-Grams
- Column Features
- Parse Features
- Regular Expressions
- Stretchy Patterns

Configure Basic Features

- Unigrams
- Bigrams
- Trigrams
- POS Bigrams
- Word/POS Pairs
- Line Length
- Contains Non-Stopwords
- Count Occurrences
- Include Punctuation
- Remove Stopwords
- Stem N-Grams

Extract

Name: features1

Rare Threshold: 5

Feature Table:

features

FEATURE_TABLE

Documents: sentiment_sentences_cv.
 Feature Plugins:
 Feature Table: features

Evaluations to Display:

Target: neg

Basic Table Statistics

- Correlation
- F-Score
- Kappa
- Precision
- Recall
- Target Hits
- Total Hits

Features in Table:

Search:

Feature

'60s
'70s
'd
'em
'll
'm
're
's
've
--

Feature Tables:

features [Save] [X]

FEATURE_TABLE

- Documents: sentiment_sentences_cv.
- Feature Plugins:
- Feature Table: features

Learning Plugin:

- Naive Bayes
- Logistic Regression
- Linear Regression
- Support Vector Machines
- Decision Trees
- Weka (All)

- Cross-Validation
- Supplied Test Set
- No Evaluation

Fold Assignment:

- Random
- By Annotation: [Dropdown]
- By File

Number of Folds:

- Auto
- Manual: 10

2 5 10 Max

Configure Weka (All)

Choose SMO -C 1.0 -L 0.001 -P 1.0

Train

Name: weka1 [Feature Selection] #: 1000

Trained Models:

weka [Save] [X]

TRAINED_MODEL

- Documents: sentiment_sentences
- Feature Plugins:
- Feature Table: features
- Learning Plugin: Weka (All)
- Wrapper Plugins:

Model Evaluation Metrics:

Metric	Value
Accuracy	0.7343
Kappa	0.4685

Model Confusion Matrix:

Act \ Pred	neg	pos
neg	2905	941
pos	1095	2721

LightSide

Extract Features | Restructure Data | Build Models | Explore Results | Compare Models | Predict Labels

Highlight: weka1

Cell Highlight:

Act \ Pred	neg	pos
neg	1149	336
pos	412	1103

Evaluations to Display:

- Frequency
- Horizontal Absolute Difference
- Horizontal Difference
- Vertical Absolute Difference
- Vertical Difference
- Model Analysis

Features in Table:

Feature	Frequency	Vertical Absolute Difference
<input type="radio"/> serious	1	0.0011
<input type="radio"/> seriously	1	0.0011
<input type="radio"/> thrills	1	0.0011
<input type="radio"/> unfortunately	1	0.0011
<input type="radio"/> value	1	0.0011
<input type="radio"/> virtually	1	0.0011
<input type="radio"/> witless	1	0.0011
<input checked="" type="radio"/> little	13	0.0011
<input type="radio"/> seems	6	0.0011
<input type="radio"/> us	8	0.0011
<input type="radio"/> beyond	3	0.0012
<input type="radio"/> classic	3	0.0012
<input type="radio"/> ending	3	0.0012
<input type="radio"/> except	3	0.0012
<input type="radio"/> gags	3	0.0012
<input type="radio"/> half	3	0.0012

Exploration Plugin: Documents Display

Filter documents by selected feature

Reverse document filter

Documents from selected cell only

Instance	Predicted	Actual	Text
<input checked="" type="checkbox"/> 124	neg	pos	we know the plot...
<input checked="" type="checkbox"/> 189	neg	pos	pretty good little ...
<input checked="" type="checkbox"/> 229	neg	pos	shyamalan offers...
<input checked="" type="checkbox"/> 563	neg	pos	daughter from da...
<input type="checkbox"/> 697	neg	pos	there are times ...
<input type="checkbox"/> 1414	neg	pos	reign of fire may ...
<input type="checkbox"/> 1690	neg	pos	it's excessively q...
<input type="checkbox"/> 1700	neg	pos	an original little fi...
<input type="checkbox"/> 2629	neg	pos	there's very little ...
<input type="checkbox"/> 2733	neg	pos	grown-up quibble...
<input type="checkbox"/> 2809	neg	pos	a sensitive and e...
<input type="checkbox"/> 2815	neg	pos	affectionately re...
<input type="checkbox"/> 2979	neg	pos	all the pieces fall ...

Instance 189 (Predicted neg, Actual pos)

Highlighting little feature hits

pretty good **little** movie .

Instance 229 (Predicted neg, Actual pos)

Highlighting little feature hits

shyamalan offers copious hints along the way -- myriad signs , if you will -- that beneath the familiar , funny surfa ce is a far bigger , far more meaningful story than one in which **little** green men come to earth for harvesting pur poses .

Report a Bug

0.4 GB used, 2.7 GB max



Any Questions?