Linguistic Agency: Implications for Computational Models of Language in Social Contexts

Carolyn P. Rosé
Language Technologies Institute
Human-Computer Interaction Institute
School of Computer Science

Carnegie Mellon
Bridging insights from linguistics to computation, and vice versa

Specifically **Sociolinguistics** and **Discourse Analysis**

Which modeling assumptions are not consistent with the theory?

What faulty assumptions about language in models lead to errors?

In what ways do models challenge theory?

In what ways are theories brittle or overly simplistic?

Linguistics  <->  Computation
Linguistic Agency
Linguistic Agency

SEMANTIC ROLE LABELING
Relational Module

AGENT
Submodule

INDIVIDUAL
LGs and DICs

GROUP
LGs and DICs

RECIPIENT
Submodule

INDIVIDUAL
LGs and DICs

GROUP
LGs and DICs
Linguistic Agency

The ability of speakers to make conscious choices about how they present themselves through language
Linguistic Agency

The ability of speakers to make conscious choices about how they present themselves through language
Outline

• Theories and Methods
  – Defining Linguistic Agency and how to study it

• From Methods to Challenges
  – Investigating Counter-Culture Vernaculars emerging through Linguistic Agency

• From Challenges to Design
  – How models of Linguistic Agency inform supportive interventions in online courses

• Into the Future
Outline

• **Theories and Methods**
  – Defining Linguistic Agency and how to study it

• **From Methods to Challenges**
  – Investigating Counter-Culture Vernaculars emerging through Linguistic Agency

• **From Challenges to Design**
  – How models of Linguistic Agency inform supportive interventions in online courses

• **Into the Future**
Methodology

Theory

Research Questions

Interpretation

Data

Patterns

The diagram represents a conceptual framework with the following components:

1. **Methodology**
   - Theory
   - Research Questions
   - Interpretation

2. **Data**
   - Patterns

The arrows indicate the flow of ideas or processes, suggesting a relationship between the components.

- Methodology influences Theory, which in turn leads to Research Questions and Interpretation.
- Data is linked to Patterns, indicating that data analysis might reveal patterns.
- The dashed box suggests an encompassing or foundational role of Methodology and Data, with Theory, Research Questions, and Interpretation building upon them.
What is a theory?
# A Spectrum of Research Methodologies

<table>
<thead>
<tr>
<th>Simplification/Abstraction</th>
<th>Preserving Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative/Positivism</td>
<td>Qualitative/Constructivism</td>
</tr>
</tbody>
</table>

## Variationist Sociolinguistics

- Women talk like X and Men talk like Y

## Interactional Sociolinguistics

- Behavior X is associated with femininity, so people engage in that behavior in order to associate themselves with feminine qualities
  
- How did a women enact X in this context?

Do X and Y predict Gender?
## A Spectrum of Research Methodologies

<table>
<thead>
<tr>
<th>Simplification/Abstraction</th>
<th>Preserving Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative/Positivism</td>
<td>Qualitative/Constructivism</td>
</tr>
<tr>
<td>Individuals</td>
<td>Pairs</td>
</tr>
</tbody>
</table>

* As we move towards more complex units of analysis, we either have to adopt a weaker methodology (which either threatens internal validity or requires us to weaken our claims), or we have to make stronger simplifying assumptions (threatens external validity).
Another view of Qualitative vs Quantitative

Quantitative research gives us categories that structure a discourse. They are like the threads that hold together a tapestry.

But qualitative analyses reveal the substance of the processes that are structured by the threads.

Every tapestry is different, and the beauty in the perceived pattern is subjective, but the experience is still valuable (as a challenge to over-simplification)
Linguistic Agency

• “The expressiveness of an individual appears to involve two radically different kinds of sign activity: the expression that he gives and the expression that he gives off.”
  – What he gives refers to the strategies employed in order to achieve social goals
  – What he gives off refers to aspects of language that reflect the internal state of the speaker that are beyond his control

Erving Goffman
Sociologist and Author of *Presentation of Self in Everyday Life*
Originally published in 1959
Outline

• Theories and Methods
  – Defining Linguistic Agency and how to study it

• From Methods to Challenges
  – Investigating Counter-Culture Vernaculars emerging through Linguistic Agency

• From Challenges to Design
  – How models of Linguistic Agency inform supportive interventions in online courses

• Into the Future
Studying Language Attitudes
Exploring the connection between language and identity in counter-cultural groups
Counter-Culture Vernaculars

• Across all three contexts, low SES counter cultures developed vernaculars used to obscure meaning
  – Sometimes referred to as “argots” or “jargons”
  – Kenya: Sheng (Swahili and English)
  – South Africa: Tsotsitaal (Zulu and Afrikaans)
  – The United States: Ghetto vernaculars (borrow from African American Vernacular English)

• Functions of counter-culture vernaculars
  – Usually associated with criminals, but may be used by youth to sound “cool”
  – Signal loyalty to in group
  – Create distance from out group
  – Protect in group through secrecy
Sheng

- Similarities with African American Vernacular English
  - Similar structure to Swahili, some claim that it is “simplified”
  - Fewer noun classes, “relaxed” rules of agreement

- Similar to Ghetto vernaculars in employing features that obscure meaning
Tsotsitaal

• Similar cultural associations
  – Literal meaning is “Gangster language”

• Similar structure and properties to Sheng
  – Grammar and vocabulary heavily borrows from both Afrikaans and Zulu
  – “over lexicalization”, may be more than one language

• The way you use it reflects your native language
Bordieu’s Theory of Distinction

• Discourse of legitimacy defines the discourse of power, which becomes the unmarked language choice
  – Affluent people can afford better education
  – In Kenya, educated people speak more English. Speaking more English is a way of showing off.

• The opposite of the discourse of power becomes associated with powerlessness
  – In Kenya, less educated people speak tribal languages and Swahili. Their English is structured more like their tribal language.
  – Language mixes are the best they can do. Educated English is unattainable. Sheng has become stigmatized.
Bordieu’s Theory of Distinction

• Low Socioeconomic status groups are vulnerable to falling into crime
  – *Well established gangs in the US are connected with international drug cartels*

• Once fallen into the counter-culture, creating distance between the counter culture and the hegemonic culture is a matter of survival
  – *Crime related counter-cultures need to communicate surreptitiously*
Bordieu’s Theory of Distinction

• Over time, through processes of distinction, the association between status and language choice becomes more entrenched
  – Sheng is a mixture of English, Swahili, and tribal languages. As part of its history, it was used in slums so kids could hide things from their parents.
  – A little bit of Sheng sounds cool. A lot of Sheng makes you sound like you’re hiding something.

• **Indexicality**: the social interpretation of a linguistic choice
Bordieu’s Theory of Distinction

• Connection with Social Meaning:
  – Shifts in usage of counter-culture vernaculars might give visibility into the relationships between subgroupings within counter-cultures
  – Example: In South Africa, non-accommodation in language choice is a way of being assertive/prideful
Implications of processes creating indexicality in language

Symbols are Arbitrary
Implications of processes creating indexicality in language

Symbols are Arbitrary but not Random
Challenges

• Counter-culture vernaculars are highly changeable
  
  – For example:

  • There are different versions of Sheng in different areas of Kenya
  • Sheng keeps changing over time
  • Tsotsitaal varies *substantially* depending on the speaker’s native language
American Street Gangs

• TheHoodUp.com
  o discussion board
  o Region specific sub-forums, but not gang specific

• Information in the profiles
  o username
  o location
  o gang affiliation

• Summary Statistics
  o Posts: 1.1 M
  o Users: 45k
  o Active Users: 3k
American Street Gangs

- **Crips, Bloods, Hoovers**
  - crips started in South Central LA
  - Pirus, Bloods, Hoovers from crips

- Chicago based
  - People Nation
    - vice lords, latin kings, stones
  - Folk nation
    - gangster disciples

- **Trinitarios**
  - hispanic gang based in NYC
Figure 1: Distribution of annotated users’ gang affiliations
Gang Alliances

Legend:
- **Alliance**
- Gang within alliance
- Set within gang
- Rivals
- Former
- Relationship

Diagram:
- **Trinitarios**
  - Latin Kings
  - Vice Lords

- **Sureños**
  - Bloods
    - Black P. Stones
    - People Nation

- **Norteños**
  - Crips
    - Hoover Criminals
      - Neighborhood / Rollin’ Os
      - Gangster Crips (Trays)
      - Gangster Disciples

- **Folks Nation**
A post from the discussion board

lol dat picc was on dat old skool nigga dat used to get on thki55 5ite " chromee " or 5um 5hit hke " exposed " cuzz i gue55 , and pointblanc was a mod on di55 5ite him and few othker got bkanned for mis using their powers or w / e fucc them niggas thko dey burbk bkanigin ass fake joke loc's
Gang related style

lol dat picc was on dat old skool nigga dat used to get on thki55 5ite " chromee " or 5um 5hit hke " exposed " cuzz i gue55 , and pointblanc was a mod on di55 5ite him and few othker got bkanned for mis using their powers or w / e fucc them niggas thko dey burbk bkangin ass fake joke loc's

avoids ck: probably a crip
uses 5 for s:
adds a k after h: probably a hoover hater
adds a k after b: probably a blood hater
uses crip jargon (loc, cuzz)
Graffiti Based Style Features

Graffiti
Social messages
Stylistic writing
crossing out other gangs

On the board

c  ck ckrab, ccome
ck  cc fucc, blocc
p  pk pkut, ...
h  hk whky, hkappens
b  bk bk1, bkang
e  3 3ast
s  5 5hit
c  c^ c^rime, c^uh
Identifying Gang

• Users who use style features (25%)

<table>
<thead>
<tr>
<th>Features</th>
<th>Number of Features</th>
<th>Accuracy</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graffiti Style Features</td>
<td>13</td>
<td>69.4%</td>
<td>0.523</td>
</tr>
<tr>
<td>Gang Name Features</td>
<td>15</td>
<td>73.7%</td>
<td>0.599</td>
</tr>
<tr>
<td>Unigrams</td>
<td>500 selected</td>
<td>78.7%</td>
<td>0.688</td>
</tr>
<tr>
<td>Gang Name + Style Features</td>
<td>28</td>
<td>81.7%</td>
<td>0.726</td>
</tr>
<tr>
<td>Unigrams + Gang Name + Style</td>
<td>500 selected</td>
<td>82.4%</td>
<td>0.740</td>
</tr>
</tbody>
</table>

Finding strong identifying characteristics does not mean these characteristics lead to high classification accuracy.
Map Task Example

‘Round vs Around

Finding strong identifying characteristics does not mean these characteristics lead to high classification accuracy.
Identifying Gang

- **All users**

<table>
<thead>
<tr>
<th>Features</th>
<th>Number of Features</th>
<th>Accuracy</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location and Ethnicity Mentions</td>
<td>12</td>
<td>32.0%</td>
<td>0.03</td>
</tr>
<tr>
<td>Graffiti Style Substitution Features</td>
<td>13</td>
<td>50.0%</td>
<td>0.305</td>
</tr>
<tr>
<td>Gang Name/Insult Frequency per Post</td>
<td>15</td>
<td>67.6%</td>
<td>0.571</td>
</tr>
<tr>
<td>Gang Name + Style</td>
<td>28</td>
<td>71.3%</td>
<td>0.623</td>
</tr>
<tr>
<td>Baseline: Unigrams</td>
<td>All (around 10,000)</td>
<td>73.5%</td>
<td>0.657</td>
</tr>
<tr>
<td>Unigrams + Style</td>
<td>500 selected</td>
<td>75.8%</td>
<td>0.686</td>
</tr>
<tr>
<td>Unigrams + Gang Name Frequency</td>
<td>500 selected</td>
<td>77.5%</td>
<td>0.710</td>
</tr>
<tr>
<td>Unigrams + Gang Name + Style</td>
<td>500 selected</td>
<td><strong>78.1%</strong></td>
<td><strong>0.718</strong></td>
</tr>
</tbody>
</table>
Thread Composition

Figure 1: Distribution of annotated users’ gang affiliations
Thread Composition

- 4 way distinction: Heterogeneous, Allied, Opposing, Mixed

- Thread Composition is associated with significant differences in expressed hostility

- Hostility by itself does not predict Thread Composition
  - Classification performance is random
Thread Composition

- 4 way distinction: Heterogeneous, Allied, Opposing, Mixed

- Thread Composition is associated with significant differences in style feature usage

- Style features are good predictors of thread composition, especially when dominant gang is taken into account
Multi-Domain Models (Daumé III, 2007)

Allows us to represent separately what is general across subpopulations and what is particular to subpopulations
Thread Composition

• Features that distinguish Allied from Opposing differ by domain.

Multi-domain version

Model Evaluation Metrics:

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.565</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.327</td>
</tr>
</tbody>
</table>

Model Confusion Matrix:

<table>
<thead>
<tr>
<th>Act \ Pred</th>
<th>allied</th>
<th>homogeneous</th>
<th>mixed</th>
<th>opposing</th>
</tr>
</thead>
<tbody>
<tr>
<td>allied</td>
<td>88</td>
<td>32</td>
<td>0</td>
<td>120</td>
</tr>
<tr>
<td>homogeneous</td>
<td>38</td>
<td>95</td>
<td>2</td>
<td>38</td>
</tr>
<tr>
<td>mixed</td>
<td>11</td>
<td>6</td>
<td>16</td>
<td>30</td>
</tr>
<tr>
<td>opposing</td>
<td>65</td>
<td>32</td>
<td>5</td>
<td>294</td>
</tr>
</tbody>
</table>

Model Evaluation Metrics:

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.627</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Model Confusion Matrix:

<table>
<thead>
<tr>
<th>Act \ Pred</th>
<th>allied</th>
<th>homogeneous</th>
<th>mixed</th>
<th>opposing</th>
</tr>
</thead>
<tbody>
<tr>
<td>allied</td>
<td>92</td>
<td>20</td>
<td>2</td>
<td>126</td>
</tr>
<tr>
<td>homogeneous</td>
<td>22</td>
<td>136</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>mixed</td>
<td>5</td>
<td>9</td>
<td>18</td>
<td>31</td>
</tr>
<tr>
<td>opposing</td>
<td>48</td>
<td>39</td>
<td>8</td>
<td>301</td>
</tr>
</tbody>
</table>
In the general space, we see a variety of positive style features in allied gangs and a variety of negative style features and derogatory name features in the negative space.

Feature Analysis

- Style features that distinguish Allied from Opposing differ by dominant gang
  - **General:**
    - Allied: $H^, B^, CC$
    - Opposing: Name features, PK, BK
  - **Bloods:**
    - Allied: Name features (for opposing gangs)
    - Opposing: $B^, CK, 3E, 5S, BK$
  - **Latin Kings:**
    - Allied: $3E, CK, CC, GDK, GD$
    - Opposing: Name Features (for opposing gangs), 5S
  - **Crips:**
    - Allied: $P^, C^, B^$, Name features (for opposing gangs)
    - Opposing: XO, CC, Name features (for Crips and Hoovers)
Due to the complex social structure of the Crips, we see some unusual patterns with Crips style features.
Feature Analysis

- Style features that distinguish Allied from Opposing differ by dominant gang
  - **General:**
    - Allied: $H^, B^, CC$
    - Opposing: Name features, PK, BK
  - **Bloods:**
    - Allied: Name features (for opposing gangs)
    - Opposing: $B^, CK, 3E, 5S, BK$
  - **Latin Kings:**
    - Allied: $3E, CK, CC, GDK, GD$
    - Opposing: **Name Features (for opposing gangs)** $5S$
  - **Crips:**
    - Allied: $P^, C^, B^$, Name features (for opposing gangs)
    - Opposing: XO, **CC**, Name features (for Crips and Hoovers)

When the dominant gang is in an opposing thread, we see features that unite the opposing gangs against them and them against opposing gangs and features that are positive towards them.
Feature Analysis

• Style features that distinguish Allied from Opposing differ by dominant gang
  • **General:**
    - Allied: H^, B^, CC
    - Opposing: Name features, PK, BK
  • **Bloods:**
    - Allied: **Name features (for opposing gangs)**
    - Opposing: B^, CK, 3E, 5S, BK
  • **Latin Kings:**
    - Allied: 3E, CK, CC, GDK, GD
    - Opposing: Name Features (for opposing gangs), 5S
  • **Crips:**
    - Allied: P^, C^, B^, **Name features (for opposing gangs)**
    - Opposing: XO, CC, Name features (for Crips and Hoovers)

*When the dominant gang is in an allied thread, we see style features that unite them against opposing gangs.*
Feature Analysis

- Style features that distinguish Allied from Opposing differ by dominant gang

  - **General:**
    - Allied: *yo*, movie, time, *cuz*
    - Opposing: forever, brothers, *food*

  - **Bloods:**
    - Allied: *gang*, *niggas*, time, rap
    - Opposing: *fake*, ya, work big

  - **Latin Kings:**
    - Allied: post, still, say, just
    - Opposing: that, love, doing, never

  - **Crips:**
    - Allied: *here*, *still*, *side*
    - Opposing: *from*, rollin, *facepalm*, *dis*

With unigrams, we see positive and negative words.

Words with style features don’t get high weight.
Conclusions

• Linguistic agency creates challenges: diverse and changing language patterns

• Models with high predictive accuracy may not be the ones that tell us the most about language

• Style features are effective predictors for thread composition
  – Connection between online and offline language practices
  – More accurate when combined with unigrams than unigrams alone
  – Reflects relationships between subgroups
  – Show interesting differences between subgroups
Outline

• Theories and Methods
  – Defining Linguistic Agency and how to study it

• From Methods to Challenges
  – Investigating Counter-Culture Vernaculars emerging through Linguistic Agency

• From Challenges to Design
  – How models of Linguistic Agency inform supportive interventions in online courses

• Into the Future
MOOCS: The Good the Bad and the Ugly

• Assess impact of chat participation on dropout along the way using a survival analysis
  – **Unit of analysis**: each 2 day period
  – **Dependent variable**: Drop = 1 on the last active time period (0 otherwise)
  – **Control variables**: Number of clicks on videos and number of clicks on discussion forums
  – **Independent variable**:  
    • Participation in collaborative chat reflection activities

![Survival Curves](image)

Reduction in attrition of between 40% and 70% across 3 studies
Investigation of State-of-the-Art Team based MOOCs

<table>
<thead>
<tr>
<th>Score = 0</th>
<th>Score = 20</th>
<th>Score = 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>#teams</td>
<td>96</td>
<td>58</td>
</tr>
</tbody>
</table>

• Dataset
  – Constructive Classroom Conversations (Stanford School of Education)
  – 12 week course (2 runs: Elementary and Secondary)

• Most teams fail!
  – final team score, based on final team project submission
  – Poor team performance from the beginning
Practical Impact?

• “The expressiveness of an individual appears to involve two radically different kinds of sign activity: the expression that he gives and the expression that he gives off.”

  – What he gives refers to the strategies employed in order to achieve social goals
  – What he gives off refers to aspects of language that reflect the internal state of the speaker that are beyond his control

Erving Goffman
Sociologist and Author of
*Presentation of Self in Everyday Life*
Originally published in 1959
Transactivity is a measure of knowledge co-construction

• Definition of Transactivity
  • building on an idea expressed earlier in a conversation
  • using a reasoning statement

Homozygous for both. One parent is orange and the other is not. Orange is dominant.

I agree because all the kids are orange also.

Self-oriented vs other-oriented
Representational vs transformational
Consensus-oriented vs Conflict-oriented
## Power, Relationships, and Transactivity

<table>
<thead>
<tr>
<th>Piaget</th>
<th>Berkowitz &amp; Gibbs</th>
<th>Kruger &amp; Tomasello</th>
<th>Azmitia &amp; Montgomery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power, Cognitive Conflict, And Learning</td>
<td>Socio-Cognitive Conflict and Transactivity</td>
<td>Power Balance And Transactivity</td>
<td>Friendship, Transactivity, And Learning</td>
</tr>
</tbody>
</table>
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)
Viewing Social Relationships through Discourse Relations

Findings from Sociolinguistics → DBM Model → Model connecting speech style accommodation and Transactivity

Reflecting Perspectives And Relationships

Findings from Developmental Psychology

Reflecting Evidence of Consensus Building
The Symmetric Accommodated Style Dependence Model

A speaker’s style at a time point depends both on mutual accommodation and the partner’s style in the previous turn.

Speech style accommodation is persistent and operates over sets of features that define latent style.

Correlation between Accommodation and Prevalence of Transactivity

$R = 0.37$
Motivation for Deliberation-based Team Formation

**Problem**

- Students have trouble finding appropriate team mates
- Teams lose critical mass through attrition
- Students depend too much on their team, lose community benefits

**Solution**

- Use community engagement as evidence of who would work well together
- Form teams later, after community engagement has started
- Maintain community connection and team connection simultaneously
Validation Study

Amazon Mechanical Turk

Diagram showing a workflow with stages labeled Preparation and Pre-task, leading to a Deliberation stage where proposals are presented. The proposals are for 'TurkerGirl's Proposal,' 'MoTrey's Proposal,' and 'Stoked's Proposal.' The diagram also indicates collaboration between Group A and Group B. A notification indicates the next step in 10 minutes.
<table>
<thead>
<tr>
<th>User</th>
<th>Time</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jay</td>
<td>10:38 AM 2/23/16</td>
<td>I believe wind power to be the best choice for the city. It will generate absolutely zero waste, it will not pollute the environment and thus will help preserve the bird and fish habitats, it produces no greenhouse gasses, and is a stable energy source.</td>
</tr>
<tr>
<td>Bob</td>
<td>10:39 AM 2/23/16</td>
<td>I would say that it is not stable because wind is not a stable thing. You can't predict when and how much you will get. Also wind turbines could present a danger to birds.</td>
</tr>
<tr>
<td>Shiny</td>
<td>10:39 AM 2/23/16</td>
<td>It would highly interrupt the bird migration patterns and thus interfere with their habitats.</td>
</tr>
<tr>
<td>Jenn</td>
<td>10:45 AM 2/23/16</td>
<td>I agree with your proposal, I think this is the best plan for the city.</td>
</tr>
<tr>
<td>kbright</td>
<td>10:46 AM 2/23/16</td>
<td>I don't think that this is a good source of energy for this town because its not reliable and a wind farm will be an eye sore to tourists. Also, it will ruin the scenery of the town.</td>
</tr>
<tr>
<td>Jaded</td>
<td>10:48 AM 2/23/16</td>
<td>This would definitely be the best choice for the environment. The only issue I see is that it isn't as reliable as the other options. It seems best used to supplement other systems.</td>
</tr>
<tr>
<td>gofer</td>
<td>10:51 AM 2/23/16</td>
<td>What steps will you take to protect the bird population from wind turbines?</td>
</tr>
<tr>
<td>kbright</td>
<td>10:51 AM 2/23/16</td>
<td>Its actually not a stable energy source. No wind no energy...</td>
</tr>
</tbody>
</table>
Transactive Deliberation Example

- **Student 1:** I think solar energy would be ideal for City A, especially if it is in the sunbelt.
- **Student 2:** I don't recall seeing solar energy as an alternative, but good idea.
- **Student 3:** For this particular city, solar energy seems like a perfectly acceptable option.
- **Student 4:** Solar energy was not listed or proposed as an alternative option, and would only be so if City A had sufficient consistent sunlight. We don't have any information proving it would be viable now, though.

A transactive reply displays reasoning and builds on an earlier contribution in the discussion.
Transactivity Based Matching

• Objective measure: Average Within group Pairwise Transactivity (AWPT)
  – Pairwise Transactivity Score (PTS): Count # transactive exchanges per pair
  – AWPT of a group is the average of PTS for all pairs within the group

• Approximate constraint satisfaction algorithm assigns students to groups of 4 such that:
  – Within groups all 4 jigsaw roles are present
  – Average AWPT over all groups in the whole population is maximized
Transactivity Based Matching

Team Collaboration

Etherpad

---

Discussion

**Sree**
I am very into plan Z, although I overall preferred nuclear energy. It's the most cost-effective without resorting to coal energy.

**gaurav**
yes

**BazaarAgent**
Hey Yohan, Can you evaluate Sree's plan from your perspective of environmental friendliness and reliability?

**Yohan**
I like the plan Sree has suggested, because wind doesn't emit greenhouse gasses and is thus environmentally friendly. But wind is not reliable.

**Xu**
Hi guys, I think plan 2 is the way to go. Cheapest plan that utilizes green tech. I see someone else suggested plan 4 and I agree with their points, it might be worth the slightly increased cost over plan 2

**BazaarAgent**
**Results**

*RQ1.* Will exposure to large community discussions lead to more successful small group collaborations?

*RQ2.* Can evidence of transactive discussions during deliberation inform the formation of more successful teams?

**Experiment 1: Group Transition Timing**
- **Preparation**: Reading & Quiz
- **Pre-task**: Proposal
- **Deliberation**: Small Group
- **Collaboration**: Move into groups after deliberation

**Experiment 2: Grouping Criteria**
- **Preparation**: Reading & Quiz
- **Pre-task**: Proposal
- **Deliberation**: Community
- **Collaboration**: Transactivity Maximization Group

**Recommendations**
- Teams should not be formed until after the students have received feedback from the course community on their individual contribution to the team project.
- Teams should be formed based on the nature of observed interaction between students during that feedback (maximizing AWPT).
Upcoming External Validation

Rise of the Superhero

• All students get basic instruction on the genre of a superhero story and the eras of superheroes
• Progression follows the course flow
  – Students will get individual instruction on superheroes in a specific period of time (Basis for jigsaw for teams in week 1)
  – Students delve into 1 period of time and design a superhero for their period of time in week 2
  – In a community pre-teaming activity, students will critique the designs of others based on genre principles and era principles in week 3
  – Teams will develop a story line for the interaction between superheroes in a new period of time (time travel narrative) in week 4 and 5
Conclusions

• Insights from Sociolinguistics and Discourse analysis enable effective model development
  – Unsupervised approach to measurement of speech style accommodation
  – Detection of Transactivity

• Understanding of underlying mechanisms informed intervention design
Outline

• Theories and Methods
  – Defining Linguistic Agency and how to study it

• From Methods to Challenges
  – Investigating Counter-Culture Vernaculars emerging through Linguistic Agency

• From Challenges to Design
  – How models of Linguistic Agency inform supportive interventions in online courses

• Into the Future
DANCE: Discussion Affordances for Natural Collaborative Exchange

About DANCE

Drawing from two decades of research in Computer Supported Collaborative Learning, we are working to design an extension of the edX platform to enhance instructionally beneficial discussion opportunities available to students. With this working group, we want to bring together people from academia and industry to build a common vision regarding what kinds of research would be valuable to the community once such a platform extension was in place to support it. Our work is initially focusing on the edX platform in particular, but in the long run we seek to provide these capabilities to Massive Open Online Courses and other online learning platforms more generally. In particular, this working group is partnering with edX as a satellite collaborative, seeking to involve researchers and developers from multiple universities, foundations, and industrial organizations.

Our foundational work is beginning with specific interventions designed to offer synchronous collaboration activities supported by intelligent conversational agents and enhancements to threaded discussions to support more intensive help exchange by leveraging social recommendation technology. However, our goals are much broader than this, seeking to leverage insights and methodologies from the field of Human-Computer Interaction and encompassing both synchronous and asynchronous communication very broadly. Our vision includes text, speech, and video based interactions, instrumented with all sorts of intelligent support powered by state-of-the-art analytics and leveraging language technologies and artificial intelligence more broadly in order to offer contextually appropriate support. We will coordinate this effort with regular online meetings and occasional in-person workshops.

Director: Carolyn Penstein Rosé
Carnegie Mellon University
cprose@cs.cmu.edu

Co-Director: Oliver Ferschke
Carnegie Mellon University
ferschke@cs.cmu.edu
Thank You!

Questions?

https://docs.google.com/forms/d/1v> 1VM/viewform?usp=send_form
Model variables

**Speech Observation**(O): Obtained from speech feature such as pitch, loudness, voice probability, harmonic to noise ratio, voice quality. We used openSMILE (Eyben et al., 2010) to extract these features.

We use BNT (Murphy, 2001) to learn parameters for our DBN models.