Brick by Brick: Iterating Interventions to Bridge the Achievement Gap with Virtual Peers

Emilee Rader, Margaret Echelbarger
Department of Communication Studies
Northwestern University
{ejrader, m-echelbarger}@northwestern.edu

Justine Cassell
Human Computer Interaction Institute
Carnegie Mellon University
justine@cs.cmu.edu

ABSTRACT

We lay out one strand of a continuing investigation into the development of a virtual peer to help children learn to use "school English" and "school-ratified science talk". In this paper we describe a detailed analysis of a corpus of child-child language use, and report our findings on the ways children shift dialects and ways of discussing science depending on the social context and task. We discuss the implications of these results for the re-design of a virtual peer that can evoke language behaviors associated with student achievement. Furthermore, our results allow us to describe the ways in which this virtual agent can tailor its level of interaction based on a child's current aptitude in this area.

Author Keywords

embodied conversational agent, virtual peer, African American English, dialect, quantitative methods

ACM Classification Keywords

H.5.1 Information Interfaces and Presentation: Multimedia Information Systems

INTRODUCTION

Significant work in the areas of pedagogical agents and cognitive tutors employs humanoid representations of teachers and tutors to help children learn science and mathematics content in schools. In this study we motivate the use of a less common technology—the *virtual peer*[11, 2]—and describe ways in which virtual peers can be designed from child-child behavioral data to scaffold children's learning of mainstream dialects and scientific literacy. The project is motivated by the persistent achievement gap between African American and European American students in the American school system [10]. Language plays an important role in this achievement gap, and children who come to school speaking African American English (AAE) rather than Mainstream American English (MAE) score lower on the GORT-3 standardized reading test independent of the socioeconomic status of the child [4], and may be unfairly judged as less skilled at using classroom discourse [8]. Our earlier work implemented a race-ambiguous virtual peer (to avoid evoking any inadvertent noxious stereotypes of ethnicity) [1], and laid out a Markov model dialogue system infrastructure that selects appropriate response for the virtual peer to enunciate based on the utterances in the child-child corpus [3]. In this paper we describe a detailed analysis of a corpus of child-child interaction, and the places in which child-child play actually results in children producing on-target school language and scientific literacy content, and discuss how this analysis can be integrated into the interaction model for the virtual peer.

In addition to dialect differences, not all children come to school fluent in the particular style of scientific discourse emphasized in traditional science classrooms [7] as well as its nonverbal correlates. Therefore an intervention that teaches how to "codeswitch" among verbal and nonverbal styles in the classroom has the potential to allow AAE speaking children to overcome an achievement barrier in the classroom and make good progress in school while maintaining their connection to their home culture and language. The use of a virtual peer derives from the recognition that children may not produce MAE or SRST in the classroom in the presence of teachers. Thus, rather than developing a tutor agent, we offer children the opportunity to interact with a peer who can model, scaffold, and evoke the target language and nonverbal behaviors, in the way that children sometimes—but not reliably—do for one another. Our hope is that by employing a peer, we will avoid an "oppositional culture" between the child's home and school environments [9].

METHOD AND PROCEDURE

Participants were 30 African American 3rd-grade children between the ages of 8 and 10 who attended schools, summer or after-school programs with largely African American participation, located in a mid- or low-SES neighborhood (85-87% low income) in a large urban metropolitan area. All of the 8-10 year old children from these venues whose parents returned a consent form appear in this sample. Initial assessment was conducted of dialect uses by one white MAE- and one African American AAE-speaking experimenter by observing children engaging in different activities, with different conversational partners of different statuses who spoke either AAE or MAE. All 30 children spoke AAE in all contexts observed and were independently described by their teachers as monodialectal and "incapable" of speaking MAE. Children were paired for the data collection and completed two tasks: first collaborating to con-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
Copyright 2011 ACM 978-1-4503-0267-8/11/05...$10.00.

struct a bridge with Legos (the "Bridge Task"), and then taking turns in the role of "Student" and "Teacher" as they explained to one another how they built the bridge (the "Classroom Task"). The tasks were videotaped from 4 angles.

Language and nonverbal behaviors were transcribed and segmented into utterances, which were then annotated for 23 morphosyntactic (i.e., grammatical) and 9 phonological (i.e., sound) features from Washington and Craig's African American English Dialect Density Measure (DDM) [5], plus two features specific to the region. Two types of scientific discourse were also coded, and characterized as school-ratified scientific talk (SRST) and non-school ratified scientific talk (NRST). SRST included questions, hypotheses, testing, explanations, and talk of causation. NRST included a narrative form of scientific discourse often discounted or discouraged in the classroom. For example, the SRST utterance, "We need to build it taller," contrasts with a possible NRST equivalent, "We got to so the fishes don't get them." Cohen's Kappa was used to assess interrater reliability, and all kappas were greater than .60 indicating good agreement.

The corpus contains 5223 utterances in 3004 speaking turns, after removing utterances that were completely unintelligible. Across both the Bridge Task and the Classroom Task, 67% of the utterances are entirely MAE and 33% contain at least one AAE dialect feature. While this may seem surprising, in children of this age, DDM is typically 10-15% and rarely more than 20% of spontaneous discourse, because AAE morphosyntax and phonology are identical to MAE in many cases. 76% of utterances contain no SRST or NRST (and were primarily questions, commands, or random comments); 5% contain NRST, and 19% contain SRST. The percentage of NRST may be low because we coded dialect and science talk independently. This avoids unintentional interpretation of all science talk spoken in AAE as NRST, when it might actually be SRST. Children spoke on average 203 utterances (SD=96), with a mean length of utterance in words of 4.5 (SD=1.17). We grouped utterances into speaking turns to focus the analysis on how one child's speech turn influences the other child's production of MAE and SRST in the subsequent turn.

ANALYSIS AND RESULTS

Our analysis focused on identifying aspects of language use that consistently precede MAE and SRST in children's speech, in order to use this correlational information to generate design hypotheses about how children adapt their dialect usage to the surrounding context that can then be implemented and tested as part of the prototype virtual peer. In other words, we wanted to know what child A did in speaking turn 1 before Child B used more MAE and SRST than expected in speaking turn 2, based on dialect usage and scientific discourse in the corpus as a whole. We used this contingency analysis approach to determine predictors of MAE and SRST in the conversation of two real peers, because our implementation of virtual peers relies on the naturalistic behaviors of peer modeling and scaffolding rather than instructional techniques such as prompts and explanations.

We conducted a series of fully-categorical mixed model logistic regressions, using maximum likelihood estimation. We used two binary dependent variables: the presence of SRST in a speaking turn, and whether a majority of utterances in a speaking turn consisted entirely of MAE. The unit of analysis is the speaking turn; all predictors describe characteristics of the previous speaking turn (i.e., the speaker's partner). The dataset includes multiple observations from each child and each pair, resulting in the potential for systematic differences resulting from between-child, within-pair, and between-pair variance to bias the results. Therefore, each model also includes random effects controls to account for variance structures due to observations that are not independent.

The selection of predictors was motivated by literature from sociolinguistics and education suggesting that particular features might result in an increase in our target "classroom language" behaviors, MAE and SRST. Unfortunately, 15 dyads did not provide enough observations in each of the categories of the higher-order interactions among the predictors to satisfy the logistic regression requirement of non-zero cells. This means it is not mathematically possible for us to include all predictors in every model, and therefore we cannot determine from this dataset which predictor is "better". Instead, the questions we ask with these models concern effect size—how big is the effect of a particular combination of predictors on our target behaviors—rather than prediction. In essence, these regressions are interpreted similarly to chi-square tests, with additional controls for non-independence of the observations.

Dialect Sub-Groups

A dialect density measure (DDM = number of dialect features / number of words) [5] was calculated for each child; higher DDM values mean more AAE dialect features. Low vs. High MAE was determined a posteriori by calculating DDM for each child and breaking them into 2 groups at the median DDM. The average DDM for High MAE children was 0.077 (SD=0.024), and the DDM for Low MAE children (those using more AAE features in their speech) was 0.163 (SD=0.033).

Results demonstrate that the children adapted their dialect usage according to the roles they assumed during the Bridge and Classroom tasks, and that this adaptation presents in different ways for the dialect sub-groups. A mixed Child Dialect (between-subjects: Low vs. High MAE) x Role (within-subjects: Peer, Student, Teacher) ANOVA showed significant main effects for both Role (df=2, F=5.31, $p < .01$) and Child Dialect (df=2, F=192.55, $p < .001$). There was also a significant Child Dialect x Role interaction (df=4, F=3.56, $p < .01$). As expected, the average DDM for Low MAE children was highest, meaning they used the most AAE, in the Peer Role: 0.18 (SD=0.05), and lower when the children role-played Teachers: 0.13 (SD=0.06). However, for these children, the DDM in the Student role was also high, nearly equivalent to the Peer Role DDM: 0.17 (SD=0.08). Thus, while these children were able to adapt their dialect usage to role, they did so only when they were pretending

	Baseline Likelihood		High MAE Partner		Low MAE Partner	
	Classroom	Bridge	Classroom	Bridge	Classroom	Bridge
High MAE Child	82%*	76%***	81%*	77%	85%*	74%.
Low MAE Child	60%	63%	76%*	62%	58%*	63% @

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 1; @ = intercept

Figure 1: Model results showing the likelihood of MAE when High vs. Low MAE-speaking children interacted with High vs. Low MAE-speaking partners, compared with the baseline model.

to be Teachers, not Students. In contrast, the High MAE children adjusted to use more MAE in both the Student and Teacher Roles.

Most importantly, Figure 1 shows that Low MAE children actually increased their use of MAE when they interacted with High MAE partners. High MAE children are likely to use the same amount of MAE in a speaking turn no matter whether the previous turn was spoken by a High or Low MAE partner—about 80-85%. In contrast, Low MAE children have only a 58% likelihood of using MAE when the previous turn was spoken by a Low MAE partner and a 76% likelihood of using MAE when the previous turn was spoken by a High MAE partner. This is important because it provides evidence that Low MAE children are sensitive to just the sort of intervention we are creating.

Turn Sequence Patterns

Additional regressions examined more closely communication patterns across speaking turns, and interactions with amount of dialect used. In describing the magnitude of the effects, we compare the predicted probabilities generated for each category of the regression models against baseline predicted probabilities from models including just dialect subgroup membership as a predictor, and the random effects controls. The predicted probabilities resulting from these baseline models can be found in the "Baseline" columns of Figures 2 and 3.

MAE as a Predictor

For High MAE children, MAE in the previous turn leads to an increased likelihood of MAE in the Bridge Task, but not the Classroom Task. These children already use significant MAE—they are already the most capable of producing school English. This result indicates that even when they are not trying to behave in a socially-acceptable fashion for a classroom setting, the presence of MAE in the previous turn influences them to use more MAE. However, for these children, MAE does NOT also lead to an increase in SRST; it actually leads to a lower-than-expected likelihood of SRST, in the Bridge Task. For Low MAE children, the presence of MAE in the previous turn leads to an increased likelihood of MAE in the next turn in the Classroom Task, but that effect is not present in the Bridge Task. Interestingly, this is the opposite pattern of the High MAE children, indicating a different process is at work for the Low MAE children. Speaking in MAE may be more difficult for them, and therefore it requires the concerted effort of role playing in the Classroom Task for them to be influenced by MAE spoken by the other child in the previous turn. For Low MAE children, MAE has no effect on the likelihood of SRST. (See Figure 2.)

		Baseline Likelihood		MAE Present		MAE Absent	
		Classroom	Bridge	Classroom	Bridge	Classroom	Bridge
DV=MAE	*High MAE Child*	82%*	76%***	82%.	78%*	80%**	65%*
	Low MAE Child	60%	63%	63%.	64%	50%*	62% @
DV= SRST	*High MAE Child*	37%***	19%	37%	17%.	36%	30%.
	Low MAE Child	29%	18%	29%	17%	28%	19% @

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 1; @ = intercept

Figure 2: Results from two models using MAE as predictor; dependent variables are the likelihood of MAE and SRST in a speaking turn.

Science Talk as a Predictor

For High MAE children, SRST in the previous turn is not associated with an increased likelihood of MAE in the next turn over what is expected. Interestingly, however, High MAE children are LESS likely than expected to produce MAE if the turn is preceded by NRST, in both the Classroom and Bridge Tasks. This seems to indicate that perhaps NRST is somehow incompatible with MAE. If a speaking turn is preceded by SRST, High MAE children are more likely to use SRST in the following turn in both the Bridge and Classroom Tasks. Like the High MAE children, MAE speech is less likely if it follows NRST in both Tasks. However, contrary to the High MAE children, both NRST and SRST lead to an increased likelihood of SRST in the Classroom Task. This hints that Low MAE children are more able to switch from NRST to SRST than the High MAE children. (See Figure 3.)

DISCUSSION AND FUTURE WORK

In their most basic form, our results demonstrate that children described by their teachers as monodialectal do have the ability to switch dialects and speak MAE. In addition, those same children, as young as 8-10 years old, are conscious of the social contexts in which each dialect is most appropriate. Finally, these children are capable of using more school-ratified science talk than they are given credit for when they are assuming the role of student and teacher. The irony of this is not lost on us: children who do not speak MAE and SRST in their actual physical classrooms, do use this dialect and style of talk when *playing* classroom.

In looking more concretely at precursors of MAE and SRST—the dialect and style of talk expected in the classroom—we find that the children who speak the least MAE in their interactions are most sensitive to the dialect of their interlocutor, indicating that the behavior of the virtual peer may be able to influence the talk of exactly the children who need this influence. There is also an interaction between the dialect density of the child and the task type. This is particularly true for low MAE speakers in the classroom task, where they are less likely to use MAE than the high MAE children in general, but more sensitive to the use of MAE in the previous turn. These low MAE children are also sensitive to the presence of any kind of science talk in the previous turn—either NRST or SRST in the previous turn leads them to produce SRST. Interestingly, for all of the children whose talk we analyzed, the presence of NRST *dampened* the likelihood of MAE, as

		Baseline Likelihood		No ScienceTalk		NRST Present		SRST Present	
		Classroom	Bridge	Classroom	Bridge	Classroom	Bridge	Classroom	Bridge
DV = MAE	*High MAE Child*	82%*	76%***	81%	77%***	74%	70%	85%	73%
	Low MAE Child	60%	63%	64%	64% @	35%	38%*	52%	59%
DV = SRST	*High MAE Child*	37%***	19%	33%	18%	37%	14%	50%	38%
	Low MAE Child	29%	18%	25%***	16% @	48%	22%	44%	30%***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 1; @ = intercept

Figure 3: Results from two models using SRST as predictor; dependent variables are the likelihood of MAE and SRST in a speaking turn.

if it were difficult to use classroom English when engaging in a narrative style of science talk.

How do the results above relate to the development of the virtual peer? It is clear, first of all, that the choice of a peer as a pedagogical tool is justified by these results since it is clear that peers can have an effect on the amount of Mainstream American English and School-Ratified Science Talk of their peers. More challengingly, however, these results indicate problems with the previous implementation of the virtual peer—where the Markov model generated the most likely utterance act to follow a given utterance act in the corpus. This is clearly by no means sufficient to elicit significant changes in the target school behaviors, in particular the increase in the kind of scientific reasoning that teachers prize. In the next iteration, the virtual peer needs to pointedly produce SRST and MAE, particularly in the classroom task. It is also clear, and technically challenging to the current implementation, that the virtual peer may need to know whether the child is already a significant user of MAE or not, since this has an effect on the kinds of influence previous turns have on MAE and SRST in subsequent turns. In general, however, we can see that to implement an algorithm into a virtual peer such that it can adapt to any child, we need to consider what our target behaviors are and what predicts them in different contexts. Our goal is to scaffold both MAE and SRST so that children who are perceived to be underachieving in the sciences, and in school in general, can utilize the virtual peer to model the appropriate behaviors.

With these results in hand we are ready to return to a redesign of the virtual peer. Previous work has demonstrated the success of using human tutoring behavior in the design of Computer Cognitive Tutors [6]. This work, for the first time, shows how data from child-child interaction can be used to improve a virtual peer to scaffold learning. The next step is to address the technological issues that will allow the virtual peer to produce these behaviors, and to test whether practice with the virtual peer transfers to actual performance in the classroom. In the meantime, our results constitute a concrete set of recommendations for a virtual peer or any classroom system targeting the use of Mainstream American English and School-Ratified Science Talk.

ACKNOWLEDGMENTS

Thanks to the RAs who helped transcribe and code; and to Lindsay Fullerton, Kathleen Geraghty, Berto Gonzalez and Rick Wash. We are grateful to NSF for generous funding under NSF ALT Award 0735664.

REFERENCES

1. J. Cassell. Culture as Social Practice: Becoming Enculturated in Human-Computer Interaction. In C. Stephanidis, editor, *Universal Access in HCI, Part III*, pages 303–313. Springer-Verlag, 2009.
2. J. Cassell, M. Ananny, A. Basu, T. Bickmore, P. Chong, D. Mellis, K. Ryokai, J. Smith, H. Vilhjálmsson, and H. Yan. Shared reality: physical collaboration with a virtual peer. In *CHI '00*, pages 259–260. 2000.
3. J. Cassell, K. Geraghty, B. Gonzalez, and J. Borland. Modeling culturally authentic style shifting with virtual peers. In *ICMI-MLMI '09*, 2009.
4. H. K. Craig, C. A. Thompson, J. A. Washington, and S. L. Potter. Performance of elementary grade African American students on the Gray Oral Reading Tests. *Language, Speech, and Hearing Services in Schools*, 35:141–154, 2004.
5. H. K. Craig and J. A. Washington. *Malik Goes to School: Examining the Language Skills of African American Students from Preschool-5th Grade*. Lawrence Erlbaum Associates, Mahwah, NJ, 2005.
6. A. C. Graesser, K. Wiemer-Hastings, P. Wiemer-Hastings, R. Kreuz, and T. T. R. Group. AutoTutor: A simulation of human tutors. *Cognitive Systems Research*, pages 35–51, 1999.
7. L. A. Kurth, R. Kidd, R. Gardner, and E. L. Smith. Student use of narrative and paradigmatic forms of talk in elementary science conversations. *Journal of Research in Science Teaching*, 39(9):26, 2002.
8. S. T. Lubienski. Are the NCTM "Standards" Reaching All Students? An Examination of Race, Class, and Instructional Practices. In *AERA Annual Meeting*, page 22, 2001.
9. J. U. Ogbu. *Black American students in an affluent suburb: a study of academic disengagement*. L. Erlbaum, Mahwah N J, 2003.
10. M. Perie, W. Grigg, and P. Donahue. The Nation's Report Card: Reading 2005. US Dept. of Educ., 2005.
11. K. Ryokai, C. Vaucelle, and J. Cassell. Virtual peers as partners in storytelling and literacy learning. *Journal of Computer Assisted Learning*, 19(2):195–208, 2003.