Knowledge Co-construction and Initiative in Peer Learning Interactions [1]

Cynthia KERSEY [a], Barbara DI EUGENIO [a], Pamela JORDAN [b] and Sandra KATZ [b]

[a] *Department of Computer Science, University of Illinois at Chicago*
[b] *Learning Research and Development Center, University of Pittsburgh*

Abstract. The aim of the project we discuss in this paper is to develop a computational model of peer learning. We present an extensive analysis of peer learning dialogues, analysis on which our computational model is based. Our model incorporates shifts of initiative as an identifier of knowledge co-construction. We have embedded this model in a peer-learning agent that collaborates with students to solve problems in the domain of computer science data structures

Keywords. Peer Learning Agent, Knowledge Co-construction, Initiative

Introduction

Peer learning has been shown to be an effective mode of learning for all participants [7,3,13]. While collaboration can be unsuccessful when members refuse to contribute or dominate the interaction [1], groups working together cooperatively are able to arrive at solutions that none could come up with individually.

The study of peer learning from a computational perspective is still in the early stages. Although some researchers have attempted to develop simulated peers ([5,14]), there is very little research on what constitutes effective peer interaction to guide the development of effective peer learning agents. We have developed KSC-PaL, an artificial agent that can collaborate with a human student via natural-language dialogue and actions within a graphical workspace. To endow KSC-PaL with appropriate behaviors, we have undertook an extensive corpus analysis in order to identify correlates of *Knowledge Co-construction* [10]. This construct explains the effectiveness of peer learning by postulating that learning is enhanced when students work together to construct knowledge.[2] However, Knowledge Co-construction (KCC) per se is too high-level a concept for a computational model, since it doesn't provide insights into what happens within any such episode. [10] extends the analysis of KCC by incorporating relations such

[1]This work is supported by the NSF (ALT-0536968) and in part by the ONR (N000140010640).

[2]Knowledge co-construction embodies a constructivist perspective: being active in the learning process, as opposed to merely listening to an explanation, is important for learning.

as *elaborate* and *criticize* within KCC episodes. Our attempts to code for similar notions were only moderately successful; but more importantly, there were no correlations between these relations and learning, and in addition, such relations are very difficult for an automatic interpreter to recognize. Hence, we looked for simpler but principled correlates of KCC. We found those in the linguistically motivated notion of *initiative shifts* in dialogue. Informally, our hypothesis is that frequent transfer of initiative in dialogue between participants indicates that they are working together to solve the problem, and hence, that they are co-constructing the solution.

Our analysis of a corpus of peer interactions confirmed this hypothesis. We found a strong relationship between initiative shifts and KCC episodes. Additionally, we found moderate correlations of learning with both KCC and with initiative shifts. This paper presents the details of this analysis along with the model derived from this analysis. We start by situating our analysis within the general goals of our project.

1. KSC-PaL

KSC-PaL [12] is an innovative peer learning agent designed to collaborate with a student to solve problems in the domain of computer science data structures. It differs from other collaborative learning agents in that it acts as a peer and can vary its behavior from more experienced peer to less experienced peer in order to encourage learning. The core of KSC-PaL is the TuTalk system which supports natural language dialogues for educational applications [11]. In developing the agent we added a graphical user interface, replaced TuTalk's student model and are adding a planner module to implement the model discussed below.

The user interface we developed manages communication between a student and KSC-PaL. It consists of a chat facility that allows the student to communicate with the agent using typed natural language dialogue, similar to an instant messaging application, and a shared graphical workspace in which the student and the agent can communicate by drawing data structures and making changes to code.

Our student model estimates student knowledge in order to provide the planner with information regarding the student's knowledge . We implemented the model using problem solution graphs [8] as Bayesian networks where each node represents either an action required to solve the problem or a knowledge concept required as part of problem solving. A user's utterances and actions are then matched to these nodes.

2. Corpus Analysis

Our corpus consists of peer learning interactions that were collected using a computer-mediated interface, identical to the one used by KSC-PaL, which allowed for communication both via text and graphical actions. These episodes are between dyads of students solving program comprehension and error diagnosis problems involving lists, stacks and binary search trees. An excerpt of an interaction is shown in the transcript sample (Figure 1). Interactions were collected for a total of 15 dyads where each dyad was presented with five problems. The problem solving session, including a pre-test and a post-test, could not exceed three hours. Therefore, not all dyads completed all five problems. The full interaction between a dyad is subdivided into as many dialogues as problems that they solved. Thus, we collected a total of 73 dialogues. In order to measure learning gains, we also presented each student with a pre-test prior to problem solving and an identical post-test at the conclusion of problem solving.

```
15:57:04 C: and printing seems to be fine
15:57:34 R: um, does it increment correctly?
15:57:46 C: yeah
15:58:01 R: we don't change where head is.
15:58:13 C: we just move p
15:58:23 C: oh wait
15:58:24 C: i see
15:58:30 C: right p = p.next
15:58:34 R: correct
```

Figure 1. A KCC episode from one of the interactions

We used multiple linear regression analysis to analyze the relationship between our target features and post-test score, which was used as an indicator of learning. Pre-test score was used as a covariate because of its significant positive correlations with post-test score (see Table 1). Separate regressions were run for each of the problem types: list (problems 2 and 3), stack (problem 4) and trees (problem 5). Problem 1 (15 dialogues) was excluded from the analysis since its purpose was to help the participants become familiar with the interface.

Table 1. Pre-test as a Predictor of Post-test

	β	R^2	p
All problems	0.88	0.77	0.00
Linked list problems	0.60	0.36	0.00
Stack problems	0.81	0.66	0.00
Tree problems	0.81	0.66	0.00

2.1. Knowledge Co-construction

In order to examine the impact of KCC on learning, the dialogues were annotated for KCC episodes. A KCC episode is defined as a series of utterances and graphical actions in which students are jointly constructing a shared meaning of a concept required for problem solving [10]. This may proceed in a variety of ways, but could include a student elaborating on a partner's contribution or criticizing a partner. Using this definition, an outside annotator and one of the authors coded

30 dialogues (approximately 46% of the corpus) for KCC episodes. The resulting intercoder reliability, measured with the Kappa statistic[4], is considered excellent ($\kappa = 0.80$). The dialogues were also annotated for *critical co-construction* where a student critically evaluates her peer's input and *elaborative co-construction* in which a student adds additional information to the topic under discussion [10]. We achieved a moderate level of intercoder reliability ($\kappa = 0.64$).

Analysis proceeded on two fronts. First, the corpus was analyzed for the relationship between KCC and learning. This was followed by an analysis of the relationship between KCC and initiative shifts.

2.1.1. Knowledge Co-construction and Learning

In order to study the relationship between KCC and learning, we correlated pre-test score plus a measure of KCC with post-test score. *KCC actions* is the number of utterances and graphical actions that occur during KCC episodes.

We found no correlations with learning in the stack or tree problems (see section 2.3). Correlations in the list problems, however, showed that being active in the learning process has a positive impact on learning for both the individual and the dyad.

In table 2, the first row shows the benefit for the dyad overall by correlating the mean post-test score with the mean pre-test score and the dyad's KCC actions. The second row shows the benefit for individuals by correlating individual post-test scores with individual pre-test scores and the dyad's KCC actions. The difference in the strength of these correlations suggests that members of the dyads are not benefitting equally from KCC. If the subjects are divided into two groups, those with a pre-test score below the mean score and those with a pre-test score above the mean score, it can be seen that those with a low pre-test score benefit more from the KCC episodes than do those with a high pre-test score (rows 3 and 4 in Table 2).

Table 2. KCC Actions as Predictor of Post-test Score (List Problems)

	β	R^2	p
KCC actions as predictor of mean post-test score	0.43	0.14	0.02
KCC actions as predictor of individual post-test score	0.33	0.08	0.03
KCC actions as predictor of individual post-test score (low pre-test subjects, n=14)	0.61	0.37	0.03
KCC actions as predictor of individual post-test score (high pre-test subjects, n=16)	0.33	0.09	ns

Co-constructing knowledge can occur in various ways such as a student elaborating on what a partner said or criticizing a partner's contribution. To explore the relationship between KCC and learning at a deeper level, we also analyzed the dialogues for each of the two types of KCC, elaborative and critical, using both of the measures of co-construction described above. In this corpus, we found no statistically significant correlations between post-test score and criticisms (after removing the impact of pre-test score) or between post-test score and elaborations (after removing the impact of pre-test score). This is in contrast to previous work [10] that showed that both types of KCC result in learning.

2.2. Initiative

There are various definitions of initiative within the computational linguistics community. Walker and Whittaker claim that initiative encompasses both dialogue and task [15]; however, several others disagree. We follow [6] in distinguishing between dialogue initiative and task initiative, namely dialogue initiative tracks who is leading the conversation and determining the current conversational focus while task initiative tracks the leader in the development of a plan to achieve a problem solving goal. In our dialogue excerpt (Figure 1) C has both dialogue and task initiative in the first utterance, but in the next utterance R takes both types of initiative. R retains task initiative until 15:58:30 when C suggests a correction to the code. However, dialogue initiative shifts to C in 15:58:13 when C takes the conversational lead.

Two coders, one of the authors and an outside annotator, coded 24 dialogues (1449 utterances, approximately 45% of the corpus) for both types of initiative. The resulting intercoder reliability is 0.77 for dialogue initiative annotation and 0.68 for task initiative, both of which are high enough to support tentative conclusions.

For dialogue initiative annotation, we used the well-known Walker and Whittaker utterance based allocation of control rules [15]. In this scheme, each utterance is tagged with one of four dialogue acts (assertion, command, question or prompt) and control is then allocated based on a set of rules.

We derived an annotation scheme for task initiative based on other research in the area [9,6]. We define task initiative as *any action by a participant to either achieve a goal directly, decompose a goal or reformulate a goal.* Some examples of task initiative in our domain are identifying a section of code as correct or incorrect and suggesting a correction to a section of code

2.2.1. Initiative and Learning

Using the coded corpus, we analyzed the relationship between learning and the number of utterances where a student held initiative as well as between learning and the number of times initiative shifted between the students. Intuitively, we assumed that frequent shifts of initiative would reflect students working together to solve the problem. As with KCC actions, we found no correlation of either type of initiative or shifts in initiative with learning in the stack and tree problems. However, in the list problems there was a significant correlation between post-test score (after removing the effects of pre-test scores) and the number of shifts in dialogue initiative and the number of shifts in task initiative (see Table 3). This analysis excluded two dyads whose problem solving collaboration had gone awry.

The correlation between both types of initiative shifts and learning are stronger than the correlation between utterances with initiative and learning. This supports the knowledge co-construction theory that learning occurs when *all* students are active in the problem solving activity.

2.2.2. Knowledge Co-construction and Initiative

Intuition suggests that KCC episodes involve frequent shifts in initiative, as both participants are actively participating in problem solving. To test this hypothesis,

Predictor of Post-test	β	R^2	p
Dialogue initiative shifts	0.45	0.20	0.00
Task initiative shifts	0.42	0.20	0.01
Utterances with dialogue initiative	0.38	0.12	0.02
Utterances with task initiative	0.14	0.04	*ns*

Table 3. Initiative Predictors of Post-test(List Problems)

we calculated the average initiative shifts per line during KCC episodes and the average initiative shifts per line during problem solving outside of KCC episodes for each dyad. A paired t-test was then used to verify that there is a difference between the two groups. The t-test showed no significant difference in average dialogue initiative shifts in KCC episodes compared with non-KCC problem solving. However, there is a significant difference between average task initiative shifts in KCC episodes compared with the rest of the dialogue (t(57) = 3.32, p = 0.0016). The moderate effect difference between the two groups (effect size = 0.49) shows that there is a meaningful increase in the number of task initiative shifts in KCC episodes compared with problem solving activity outside of the KCC episodes. Analyzing only the list problems shows an even stronger effect (effect size = 0.65).

We then computed a second measure of KCC: *KCC initiative shifts* is the number of task initiative shifts that occur during KCC episodes. This measure is meant to reflect the density of the KCC episodes, where many task initiative shifts reflect more active KCC.

Table 4 uses KCC initiative shifts as the measure of co-construction. It shows similar results to table 2, where KCC actions was used. It is interesting to note that when the outlier dyads were removed (see section 2.2), the correlation with learning is much stronger for the low pre-test score subjects when KCC initiative shifts are used as the measure of KCC ($R^2 = 0.45$, $p = 0.02$) than when KCC actions are used (cf. Table 4).

Table 4. KCC/Initiative as Predictors of Post-test Score (List Problems)

	β	R^2	p
KCC initiative shifts as predictor of mean post-test score	0.46	0.15	0.01
KCC initiative shifts as predictor of individual post-test score	0.35	0.09	0.02
KCC initiative shifts as predictor of individual post-test score (low pre-test subjects, n=14)	0.41	0.17	0.16
KCC initiative shifts as predictor of individual post-test score (low pre-test subjects, outliers removed n=12)	0.67	0.45	0.02
KCC initiative shifts as predictor of individual post-test score (high pre-test subjects, n=16)	0.10	0.01	ns

2.3. Discussion

For the linked list problems the corpus analysis shows that there is a correlation between knowledge co-construction and learning, as other research suggests. We also found a correlation between initiative shifts and learning which confirms the theory that peer learning is most effective when all students are active participants. Additionally we found a strong relationship between task initiative shifts and KCC episodes.

The fact that only the linked list problems showed a correlation of initiative and KCC with learning is likely caused by the variations in student ability in the different problem types. The lack of correlations in the tree problem is potentially caused by the wide variation in experience level of the students which would inhibit KCC. In the stack problem, the students had a better understanding of stacks prior to problem solving and thus less time was spent in discussion and problem solving.

3. Current and Future Work

Since the corpus analysis showed a correlation between task initiative and KCC and between task initiative and learning, the next step is to have KSC-Pal encourage initiative shifts, under appropriate conditions. We define these conditions as follows: KSC-PaL will encourage initiative shifts when (1) the knowledge score has not increased in a specified period of time and (2) the number of task initiative shifts is less than the average initiative shifts for the current problem.

As concerns (1), knowledge score is a measure derived by taking the sum of the probabilities of the nodes in the current problem's solution graph.

As concerns (2) we need to be able to recognize task initiative shifts, or the lack thereof, in real time. We explored two different methods to do so. One is that student uncertainty may lead to a shift in initiative. The other is that certain cues for initiative shifts identified in related literature[6,15] lead to initiative shifts.

Intuitively, uncertainty by a peer might lead his partner to take the initiative. One possible identifier of student uncertainty is hedging. So we annotated utterances in our peer dialogues with hedging categories as identified in Bhatt et. al [2]. Using these categories we were unable to reliably annotate for hedging. But, after collapsing the categories into a single binary value of hedging/not hedging we arrived at an acceptable agreement ($\kappa = 0.71$). Another identifier of uncertainty is a student's request for feedback from his partner. When uncertain of his contribution, a student may request an evaluation from his peer. So, we annotated utterances with "request for feedback" and were able to arrive at an excellent agreement ($\kappa = 0.82$).

We also explored cues, such as invalidity, that Chu-Carroll and Brown [6] identify as indicative of task initiative shifts.

Using a combination of these cues and student uncertainty, we were able to predict 283 shifts in task initiative or approximately 67% of all task initiative shifts. The remaining shifts were likely explicit take overs of initiative without preceding indicators.

Since we identified several ways to predict and encourage initiative shifts, we then evaluated which of these identifiers more often resulted in an initiative shift and which of these initiative shifts more often led to an increase in knowledge score. The three identifiers that were found to most often lead to a shift in task initiative and result in an increase in knowledge score are: using prompts; making a mistake which will ideally lead to a partner's criticism; and requesting feedback.

The model described above has incorporated into KSC-PaL's planner module. The planner will select scripts to encourage initiative shifts, when necessary, in order to enhance student learning. These scripts will have the agent vary its behavior from a less experienced peer that hedges and makes mistakes to a more experienced peer that assists a struggling student in problem solving.

References

[1] Brigid Barron. Problem solving in video-based microworlds: Collaborative and individual outcomes of high-achieving sixth-grade students. *Journal of Educational Psychology*, 92(2):391–398, June 2000.

[2] Khelan Bhatt, Martha Evens, and Shlomo Argamon. Hedged responses and expressions of affect in human/human and human computer tutorial interactions. In *Proceedings Cognitive Science*, 2004.

[3] A. L. Brown and A. S. Palincsar. *Guided, cooperative learning and individual knowledge acquisition*, pages 307–226. Lawrence Erlbaum Associates, Hillsdale, NJ, 1989.

[4] Jean Carletta. Assessing agreement on classification tasks: the kappa statistic. *Comput. Linguist.*, 22(2):249–254, 1996.

[5] Tak-Wai Chan and AB Baskin. Studying with the prince. In *Proceedings of the ITS-88 Conference*, pages 194–200, 1988.

[6] Jennifer Chu-Carroll and Michael K. Brown. An evidential model for tracking initiative in collaborative dialogue interactions. *User Modeling and User-Adapted Interaction*, 8(3–4):215–253, September 1998.

[7] P.A. Cohen, J.A. Kulik, and C.C. Kulik. Education outcomes of tutoring: A meta-analysis of findings. *American Education Research Journal*, 19(2):237–248, 1982.

[8] Cristina Conati, Abigail Gertner, and Kurt Vanlehn. Using bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interaction*, 12(4):371–417, 2002.

[9] Curry I. Guinn. An analysis of initiative selection in collaborative task-oriented discourse. *User Modeling and User-Adapted Interaction*, 8(3-4):255–314, 1998.

[10] Robert G.M. Hausmann, Michelene T.H. Chi, and Marguerite Roy. Learning from collaborative problem solving: An analysis of three hypothesized mechanisms. In K.D Forbus, D. Gentner, and T. Regier, editors, *26th Annual Conference of the Cognitive Science Society*, pages 547–552, Mahwah, NJ, 2004.

[11] Pamela W Jordan, Brian Hall, Michael A. Ringenberg, Yui Cue, and Carolyn Penstein Rosé. Tools for authoring a dialogue agent that participates in learning studies. In *Artificial Intelligence in Education, AIED 2007*, pages 43–50, 2007.

[12] Cynthia Kersey, Barbara Di Eugenio, Pamela W Jordan, and Sandra Katz. Modeling knowledge co-construction for peer learning interactions. In *ITS 2008, The 9th International Conference on Intelligent Tutoring Systems, Student Research Workshop*, Montreal, Canada, 2008.

[13] M. D. Rekrut. Teaching to learn: Cross-age tutoring to enhance strategy instruction. *American Education Research Association*, 1992.

[14] Aurora Vizcaíno. A simulated student can improve collaborative learning. *International Journal of Artificial Intelligence in Education*, 15(1):3–40, 2005.

[15] Marilyn Walker and Steve Whittaker. Mixed initiative in dialogue: an investigation into discourse segmentation. In *Proceedings of the 28th annual meeting on Association for Computational Linguistics*, pages 70–78, Morristown, NJ, USA, 1990. Association for Computational Linguistics.