Metacognition and Learning in Spoken Dialogue Computer Tutoring

Kate Forbes-Riley and Diane Litman

Learning Research and Development Center, University of Pittsburgh, 3939 O'Hara St., Pittsburgh, PA 15260
{forbesk,litman}@cs.pitt.edu

Abstract. We investigate whether four metacognitive metrics derived from student correctness and uncertainty values are predictive of student learning in a *fully automated* spoken dialogue computer tutoring corpus. We previously showed that these metrics predicted learning in a comparable wizarded corpus, where a human wizard performed the speech recognition and correctness and uncertainty annotation. Our results show that three of the four metacognitive metrics remain predictive of learning even in the presence of noise due to automatic speech recognition and automatic correctness and uncertainty annotation. We conclude that our results can be used to inform a future enhancement of our fully automated system to track and remediate student metacognition and thereby further improve learning.

Keywords: metacognition, learning, correlations, spoken dialogue computer tutor, automatic speech recognition and correctness and uncertainty annotation, natural language processing.

1 Introduction

Metacognition is an important area of intelligent tutoring systems research, both in and of itself and with respect to its relationship to learning (e.g. [1,2]). Within tutorial dialogue, one metacognitive state that has received a lot of interest is student uncertainty. In particular, researchers have hypothesized that student uncertainty and incorrectness both signal "learning impasses", i.e. student learning opportunities [3]. In addition, multiple correlational studies have shown a link in tutorial dialogue between learning and student uncertainty or the related state of confusion [4,5,6]. Furthermore, although most computer tutors respond based only on student correctness, a number of controlled experiments have investigated the benefits of responding to student uncertainty over and above correctness during computer tutoring [7,8,9,10,11]. Some of these experiments have shown that responding to student uncertainty over and above correctness results in improved tutoring system performance, as measured by student learning, user satisfaction, and dialogue or learning efficiency.

Drawing on the metacognition literature, we are investigating relationships between the student states of uncertainty and correctness via complex metacognitive metrics that combine measures of these two states. Other researchers

V. Aleven, J. Kay, and J. Mostow (Eds.): ITS 2010, Part I, LNCS 6094, pp. 379–388, 2010.

have previously used such metacognitive metrics to investigate multiple dimensions very similar to uncertainty and correctness, and we use and build on this literature. Our metrics include learning impasse severity [12] and knowledge monitoring accuracy [13], as well as bias (i.e., over/under confidence) and discrimination (e.g., uncertainty primarily about incorrect answers) [14]. In prior work, we investigated the relationship between these four metacognitive metrics and learning in a wizarded spoken tutoring dialogue corpus, where speech recognition and uncertainty and correctness annotation were performed in real-time by a human "wizard" [5,6]. We computed themetacognitive metrics from the wizard's annotations. We showed that although student uncertainty during the tutoring dialogues does not predict learning, average learning impasse severity, knowledge monitoring accuracy and discrimination were all predictive of learning.

In this paper, we investigate whether these metacognitive metrics remain predictive of learning in a comparable corpus that was collected using the *fully automated* version of our computer tutor. We computed two sets of metacognitive metrics: one set computed from the system's real-time automatic annotations of uncertainty and correctness, and one set computed from manual annotations of uncertainty and correctness performed after the experiment was over. Our results show that almost all of the same metacognitive metrics that predict learning during the wizarded computer tutoring also predict learning during the fully automated computer tutoring, using either the automatically-computed or manually-computed metacognitive metrics. We conclude that these metacognitive metrics are a useful construct for understanding student learning during spoken dialogue computer tutoring, even in the presence of noise introduced by fully automated student uncertainty detection and speech and natural language processing. Our results will be used to track and remediate metacognition in future system versions and thereby further improve student learning.

2 Spoken Dialogue Computer Tutoring Data

This research uses the ITSPOKE-AUTO corpus, which is a collection of dialogues between college students and our spoken dialogue computer tutor, ITSPOKE (**I**ntelligent **T**utoring **SPOKE**n dialogue system). ITSPOKE is a speech-enhanced version of the Why2-Atlas qualitative physics tutor [15].

The ITSPOKE-AUTO corpus is the second of two corpora collected over two prior controlled experiments evaluating the utility of enhancing ITSPOKE to respond to learning impasses involving student uncertainty, over and above correctness [8]. Motivated by research that views uncertainty as well as incorrectness as signals of "learning impasses" [3] (i.e., opportunities to learn), ITSPOKE was modified for use in these two experiments so that it associated one of four impasse states with every student answer, and could adapt contingently based on each answer's impasse state (in the experimental conditions), or based only on its correctness (in the control conditions). The four impasse states correspond to all possible combinations of (binary) uncertainty (uncertain (**UNC**), certain

Nominal State:	INCOR_CER	INCOR_UNC	COR_UNC	COR_CER
Severity Rank:	most (3)	less (2)	least (1)	none (0)

Fig. 1. Different Impasse State Severities

(**CER**)[1]) and correctness (incorrect (**INCOR**), correct (**COR**)), as shown in Figure 1. The incorrectness component of each state reflects the actual accuracy of the student's answer, while the uncertainty component reflects the tutor's perception of the student's awareness of this accuracy. The scalar ranking of impasse states in terms of severity combines these two components and will be discussed below. Further details of the adaptive system are discussed elsewhere [7].

For the two experiments, the experimental procedure was as follows: students (1) read a short physics text, (2) took a multiple-choice pretest, (3) worked through five problems (1 per dialogue) with a version of the system, (4) took a survey, and (5) took an isomorphic posttest.

The first corpus, called the ITSPOKE-WOZ corpus [8], contains 405 dialogues from 81 students, and was collected from the first experiment using a semi-automatic version of ITSPOKE in which speech recognition and correctness and uncertainty annotation were performed by a human "wizard", to test the upper-bound performance of adapting to uncertainty (i.e. without the noise introduced by speech and language processing). The wizard listened to the dialogues (without students' knowledge) and labeled each answer with an uncertainty (UNC, CER) and a correctness label (INCOR, COR).[2] Average pretest and posttest scores were 0.51 and 0.75, respectively.

The second corpus, called the ITSPOKE-AUTO corpus, contains 360 dialogues from 72 students, and was collected from the second experiment using a fully automated version of ITSPOKE in which all tasks were fully automated. In particular, student speech was digitized from microphone input and sent to the Sphinx2 speech recognizer whose stochastic language models were trained on the ITSPOKE-WOZ corpus and prior ITSPOKE corpora. Correctness was automatically labeled on the speech recognition output using the TuTalk semantic analyzer [16], which was trained on the ITSPOKE-WOZ corpus. Uncertainty was automatically labeled on the speech recognition output using an uncertainty model built with WEKA software from features of the student speech and dialogue context, including lexical, pitch, temporal, and energy features as well as tutor question and gender. The uncertainty model is a logistic regression equation that was trained on the ITSPOKE-WOZ corpus.

The ITSPOKE-AUTO corpus was manually transcribed after the second experiment, and the wizard from the first experiment manually labeled each student turn in the ITSPOKE-AUTO corpus for uncertainty and correctness. The system's correctness and uncertainty labels were then compared with the wizard's labels: for correctness, this comparison yielded a system accuracy of

[1] A 'certain' answer may actually be certain or neutral for certainty.

[2] In prior ITSPOKE corpora, this wizard displayed interannotator agreement of 0.85 and 0.62 Kappa on correctness and uncertainty, respectively [5].

ITSPOKE-AUTO$_1$: Let's begin by looking at the motion of the man and his keys while he's holding them. How does his velocity compare to that of his keys?

STUDENT-AUTO$_1$: his also the is the same as that of his keys points force **(INCOR_CER)**

STUDENT-MANU$_1$*: his velocity is the same as that of his keys while he's holding them (COR_UNC)*

ITSPOKE-AUTO$_2$: Actually... Their velocities must be the same, because the man's holding his keys, so they're not moving with respect to each other. [...] What are the forces exerted on the man after he releases his keys?

STUDENT-AUTO$_2$: the only force is **(INCOR_UNC)**

STUDENT-MANU$_2$*: the only force is (INCOR_UNC)*

ITSPOKE-AUTO$_3$: Well... Gravity is the only force acting on the man. [...] So what's the direction of the force of gravity on the man (e.g., vertically up)?

STUDENT-AUTO$_3$: vertically down **(COR_CER)**

STUDENT-MANU$_3$*: vertically down (COR_CER)*

Fig. 2. Annotated ITSPOKE-AUTO Corpus Excerpt

84.7%, and for uncertainty, this comparison yielded a system accuracy of 80.3% Speech recognition accuracy measured as 100% - Word Error Rate (WER) was 74.6%. WER is a standard measure for evaluating the performance of automatic speech recognition software. We compute WER using the NIST sclite program, which uses minimum edit distance to come up with a WER score by assigning penalties for word insertions, deletions, and substitutions. Average pretest and posttest scores in the ITSPOKE-AUTO corpus were 0.51 and 0.73, respectively.[3]

Figure 2 shows an annotated example of the ITSPOKE-AUTO corpus dialogues. **STUDENT-AUTO** shows the automatically recognized speech and automatic uncertainty and correctness labels for each student turn, while *STUDENT-MANU* shows the corresponding manual transcript and annotations.

3 Metacognitive Performance Metrics

In this section we introduce several ways of combining uncertainty and correctness annotations into quantitative metacognitive performance metrics. All metrics were computed on a per student basis (over all five dialogues). In addition, all metrics were computed twice: once based on the automatic correctness and uncertainty annotations (*-auto*), and once based on the corresponding manual annotations (*-manu*). Finally, note that our metrics represent inferred (or tutor-perceived) values rather than actual values, because our uncertainty labeling is done by the system or a human judge; we discuss this issue in Section 5.

[3] Independent repeated measures ANOVA analyses of both corpora showed significant main effects for repeated test measure, indicating that students learned a significant amount during both experiments.

Our first metric is based on a ranking of learning impasses by severity. In particular, we first associated a scalar **impasse severity** value with each student answer in the ITSPOKE-AUTO corpus, and then computed an average impasse severity. Our impasse severity values were proposed in our earlier work [12] and are shown in Figure 1. According to our ranking, the most severe type of impasse (3) occurs when a student is incorrect but not aware of it. States 2 and 1 are of increasingly lesser severity: the student is incorrect but aware that s/he might be, and the student is correct but uncertain about it, respectively. Finally, no impasse exists when a student is correct and not uncertain about it (0). These severity rankings reflect our assumption that to resolve an impasse, a student must first perceive that it exists. Incorrectness simply indicates that the student has reached an impasse, while uncertainty - in a correct or incorrect answer - indicates that the student perceives s/he has reached an impasse.

The rest of our metacognitive metrics are taken from the metacognitive performance literature. The knowledge monitoring accuracy metric that we use is the Hamann coefficient **(HC)** [13]. This metric has previously been used to measure the accuracy of one's own knowledge monitoring, called "Feeling of Knowing"(FOK) [17]. A closely related notion in the metacognition literature is "Feeling of Another's Knowing" (FOAK), which refers to monitoring the FOK of someone else [18], and is very similar to our student uncertainty labeling as performed by the system or a human judge. High and low FOK/FOAK judgments have also been associated with speaker certainty and uncertainty, respectively, in prior research [19].

HC measures absolute knowledge monitoring accuracy[4], or the accuracy with which certainty reflects correctness. HC ranges in value from -1 (no knowledge monitoring accuracy) to 1 (perfect accuracy). As shown below, the numerator subtracts cases where (un)certainty is at odds with (in)correctness from cases where they correspond, while the denominator sums over all cases.

$$\mathbf{HC} = \frac{(COR\_CER+INCOR\_UNC)-(INCOR\_CER+COR\_UNC)}{(COR\_CER+INCOR\_UNC)+(INCOR\_CER+COR\_UNC)}$$

Following [20], who investigate the role of immediate feedback and other metacognitive scaffolds in a medical tutoring system, we additionally measure metacognitive performance in terms of **bias** and **discrimination** [14]. Bias measures the overall degree to which confidence matches correctness. Bias scores greater than and less than zero indicate overconfidence and underconfidence, respectively, with zero indicating best metacognitive performance. As shown below, the first term represents the relative proportion of confidet answers (certain cases/all cases); the second represents the relative proportion of correct answers.

$$\mathbf{bias} = \frac{COR\_CER+INCOR\_CER}{COR\_CER+INCOR\_CER+COR\_UNC+INCOR\_UNC} - \frac{COR\_CER+COR\_UNC}{COR\_CER+INCOR\_CER+COR\_UNC+INCOR\_UNC}$$

[4] While Gamma (which measures relative monitoring accuracy) is also often used, there is a lack of consensus regarding the benefits of Gamma versus HC [13], and we found HC more predictive of learning in our ITSPOKE-WOZ corpus [6].

Table 1. Prior Correlation Results from ITSPOKE-WOZ Corpus

Measure	Mean	SD	R	p
AV Impasse Severity	.63	.24	-.56	.00
HC	.59	.16	.42	.00
Bias	-.02	.12	-.21	.06
Discrimination	.42	.19	.32	.00
%C	.79	.09	.52	.00
%U	.23	.11	-.13	.24

Discrimination measures the ability to discriminate performance in terms of (in)correctness. Discrimination scores greater than zero indicate higher metacognitive performance. As shown below, the first term represents the proportion of correct answers judged as certain, and the second term represents the proportion of incorrect answers judged as certain.

$$\textbf{discrimination} = \frac{COR\_CER}{COR\_CER+COR\_UNC} - \frac{INCOR\_CER}{INCOR\_CER+INCOR\_UNC}$$

To illustrate the computation of our four metacognitive performance metrics, suppose the annotated dialogue excerpt in Figure 2 represented our entire dataset (from a single student). Then we would have the following values for our automatically-derived (*_auto*) metrics for that student:

$$AVImpasseSeverity\_auto = \frac{3+2+0}{3} = \frac{5}{3}$$
$$HC\_auto = \frac{(1+1)-(1+0)}{(1+1)+(1+0)} = \frac{1}{3}$$
$$bias\_auto = \frac{1+1}{1+1+0+1} - \frac{1+0}{1+1+0+1} = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}$$
$$discrimination\_auto = \frac{1}{1+0} - \frac{1}{1+1} = \frac{1}{1} - \frac{1}{2} = \frac{1}{2}$$

In prior work [5,6], we showed that these four metacognitive metrics were predictive of learning in our ITSPOKE-WOZ corpus, where speech recognition, and uncertainty and correctness annotation were performed by a wizard. We computed the partial Pearson's correlation between each metacognitive metric and posttest, after first controlling for pretest to account for learning gain. We also computed the correlation for the percentage of student turns manually annotated as correct (**%C**) and as uncertain (**%U**). Correctness and uncertainty are useful baselines since they were used to derive the four complex metrics and have previously been shown to predict learning by ourselves and others (e.g [21]). Table 1 summarizes the results of this prior work, showing the mean and standard deviation of each metric, along with its Pearson's Correlation Coefficient (R), and the significance of the correlation (p).

4 Results

Here we investigate whether our four metacognitive metrics are predictive of learning in our "noisy" ITSPOKE-AUTO corpus, where speech recognition, uncertainty and correctness annotation were fully automated.

Comparison of Table 2 and Table 1 shows that with the exception of discrimination, the two metacognitive metrics (impasse severity and knowledge monitoring accuracy) that are significantly correlated with learning in the ITSPOKE-WOZ corpus remain correlated with learning in the ITSPOKE-AUTO corpus, both when derived from the automatic (*_auto*) and the manual annotations (*_manu*). In the case of average impasse severity, both the automatically-derived and manually-derived metrics yield a negative correlation, but the manually-derived metric (R = -0.50) is closest to the result in the ITSPOKE-WOZ corpus (R = -0.56). In the case of knowledge monitoring accuracy (HC), both the automatically-derived and manually-derived metrics yield a positive correlation, but the automatically-derived metric (R = 0.35) is closest to the result in the ITSPOKE-WOZ corpus (R = 0.42). Bias is negatively correlated with learning as a trend in the ITSPOKE-WOZ corpus; in the ITSPOKE-AUTO corpus the manually-derived bias metric is nearly but not quite a trend while the automatically-derived bias metric is significant. These results suggest that less severe impasse states (i.e., impasses that include uncertainty), greater knowledge monitoring accuracy, and underconfidence about one's correctness, are all better for the student from a learning perspective during computer tutoring, even when the measurement of these metrics must take into account noise due to automatic uncertainty detection and natural language processing.

Interestingly, the simple uncertainty metric (**%U**) in and of itself does not show predictive utility in this data; it is not correlated with learning in the ITSPOKE-AUTO corpus, nor did it correlate with learning in the ITSPOKE-WOZ corpus. However, correctness **%C** does significantly correlate with learning in both corpora; the manually-derived metric is closer to the ITSPOKE-WOZ corpus (R = 0.52) than the automatically-derived metric (R = 0.39).

Although these results suggest remediating metacognition can have a positive impact on learning in both wizarded and fully automated spoken dialogue tutoring, they also raise the question of whether this will be effective over remediating correctness. We addressed this question via three further analyses. First we

Table 2. Correlation Results from ITSPOKE-AUTO Corpus

Metric	Mean	SD	R	p
AV Impasse Severity_auto	.96	.26	-.40	.00
AV Impasse Severity_manu	.82	.23	-.50	.00
HC_auto	.42	.14	.35	.00
HC_manu	.49	.13	.29	.02
Bias_auto	.21	.07	-.36	.00
Bias_manu	.06	.13	-.19	.11
Discrimination_auto	.19	.10	-.04	.77
Discrimination_manu	.30	.14	-.03	.81
%C_auto	.66	.10	.39	.00
%C_manu	.72	.09	.52	.00
%U_auto	.13	.07	-.15	.20
%U_manu	.22	.14	-.13	.28

computed bivariate Pearson's correlations between correctness and each metacognitive metric. Correctness was significantly correlated with all metacognitive metrics in both the ITSPOKE-WOZ and ITSPOKE-AUTO corpora (both manually and automatically-derived). This suggests that remediating megacognition will not add value over remediating correctness. However, we then computed partial Pearson's correlations between each metacognitive metric and posttest after controlling for pretest and correctness. In the ITSPOKE-WOZ corpus, all metrics except bias remained significantly correlated with posttest, but in the ITSPOKE-AUTO corpus, no metric remained correlated with posttest. Finally, we computed stepwise linear regressions that allowed the model to select from pretest, correctness and the metacognitive metrics. In the ITSPOKE-WOZ corpus HC was selected for inclusion in the regression model after %C and pretest [6]; this indicates that knowledge monitoring accuracy adds value over and above correctness for predicting learning. In the ITSPOKE-AUTO corpus, AV Impasse Severity_auto was selected besides pretest when using automatically-derived metrics, but %C_manu was selected besides pretest when using manually-derived metrics. These last two analyses suggest that remediating metacognition can add value over remediating correctness in the "ideal" and the "realistic" conditions of wizarded and fully automated spoken dialogue tutoring, respectively.

5 Conclusions and Future Directions

This paper investigates whether four metacognitive metrics remain predictive of student learning in a previously collected fully automated spoken dialogue computer tutoring corpus; we previously showed that these metacognitive metrics predict learning in a comparable wizarded corpus. Our purpose in this study was to determine whether our prior results could be replicated even in the presence of noise due to automatic speech recognition and automatic correctness and uncertainty annotation. Our larger goal is to use our results to track and remediate metacognition and thereby further improve student learning

Of our four metacognitive metrics, one was introduced in our prior work (impasse severity); the other three come from the metacognitive performance literature (knowledge monitoring accuracy, bias and discrimination). We computed one set of metacognitive metrics from the system's real-time automatic annotations of uncertainty and correctness, and another set from subsequent manual annotations. Our results show that average impasse severity, knowledge monitoring accuracy and bias remain predictive of learning in the fully automated corpus - both when computed from the automatic values and when computed form the manual values. We conclude that these metacognitive metrics are a useful construct for understanding student learning during spoken dialogue computer tutoring, even when their measurement includes noise introduced by fully automated uncertainty detection and natural language processing. Furthermore our analyses suggest that remediating metacognitve metrics can add value over and above remediating correctness; this result is strongest in the "ideal" conditions

of wizarded tutoring, but our regression results suggest that it also holds in the "realistic" conditions of fully automated spoken dialogue tutoring.

In future work we plan to use our results to inform a modification of our system aimed at improving student metacognitive abilities and also thereby improving student learning. In particular, our results indicate that it feasible to develop enhancements for our fully automated system that target student metacognition based on the noisy version of our metacognitive metrics; if our results had not held for our automatically-derived metrics then we would have to explore system enhancements that target student metacognition using the much more time-consuming and expensive wizarded system. Note however that because uncertainty in our system is labeled by the tutor (either the system or a human wizard), our metacognitive metrics represent *inferred* or tutor-perceived values rather than actual values. It is well known in the affective tutoring literature that obtaining "actual" values for student/user affective states and attitudes is difficult; for example, student self-judgments and peer judgments have both been shown to be problematic (e.g. [22]). Nevertheless, to help measure improvements in student metacognitive abilities due to our future system modifications, we will also incorporate "Feeling of Knowing" (FOK) ratings into our testing, whereby students will provide input on their uncertainty levels. More generally, there is increasing interest in using intelligent tutoring systems to teach metacognition, and we plan to build on this literature (e.g. [1,2,20]) with future system enhancements that target student metacognitive abilities.

Acknowledgments

This work is funded by National Science Foundation (NSF) award #0914615 and #0631930. We thank Art Ward for comments.

References

1. Aleven, V., Roll, I. (eds.): AIED Workshop on Metacognition and Self-Regulated Learning in Intelligent Tutoring Systems (2007)
2. Roll, I., Aleven, V. (eds.): ITS Workshop on Meta-Cognition and Self-Regulated Learning in Educational Technologies (2008)
3. VanLehn, K., Siler, S., Murray, C.: Why do only some events cause learning during human tutoring? Cognition and Instruction 21(3) (2003)
4. Craig, S., Graesser, A., Sullins, J., Gholson, B.: Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. Journal of Educational Media 29(3) (2004)
5. Litman, D., Forbes-Riley, K.: Improving (meta)cognitive tutoring by detecting and responding to uncertainty. In: Working Notes of the Cognitive and Metacognitive Educational Systems AAAI Symposium, Arlington, VA (November 2009)
6. Litman, D., Forbes-Riley, K.: Spoken tutorial dialogue and the feeling of another's knowing. In: Proceedings 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), London, UK (September 2009)

7. Forbes-Riley, K., Litman, D.: Designing and evaluating a wizarded uncertainty-adaptive spoken dialogue tutoring system. In: Computer Speech and Language, CSL (2010) (in press)
8. Forbes-Riley, K., Litman, D.: Adapting to student uncertainty improves tutoring dialogues. In: Proc. Intl. Conf. on Artificial Intelligence in Education (2009)
9. Pon-Barry, H., Schultz, K., Bratt, E.O., Clark, B., Peters, S.: Responding to student uncertainty in spoken tutorial dialogue systems. International Journal of Artificial Intelligence in Education 16, 171–194 (2006)
10. Aist, G., Kort, B., Reilly, R., Mostow, J., Picard, R.: Experimentally augmenting an intelligent tutoring system with human-supplied capabilities: Adding human-provided emotional scaffolding to an automated reading tutor that listens. In: Proc. Intelligent Tutoring Systems Workshop on Empirical Methods for Tutorial Dialogue Systems (2002)
11. Tsukahara, W., Ward, N.: Responding to subtle, fleeting changes in the user's internal state. In: Proc. SIG-CHI on Human Factors in Computing Systems (2001)
12. Forbes-Riley, K., Litman, D., Rotaru, M.: Responding to student uncertainty during computer tutoring: A preliminary evaluation. In: Woolf, B.P., Aïmeur, E., Nkambou, R., Lajoie, S. (eds.) ITS 2008. LNCS, vol. 5091, pp. 60–69. Springer, Heidelberg (2008)
13. Nietfeld, J.L., Enders, C.K., Schraw, G.: A monte carlo comparison of measures of relative and absolute monitoring accuracy. Educational and Psychological Measurement (2006)
14. Kelemen, W.L., Frost, P.J., Weaver, C.A.: Individual differences in metacognition: Evidence against a general metacognitive ability. Memory and Cognition 28, 92–107 (2000)
15. VanLehn, K., Jordan, P.W., Rosé, C., Bhembe, D., Böttner, M., Gaydos, A., Makatchev, M., Pappuswamy, U., Ringenberg, M., Roque, A., Siler, S., Srivastava, R., Wilson, R.: The architecture of Why2-Atlas: A coach for qualitative physics essay writing. In: Cerri, S.A., Gouardéres, G., Paraguaçu, F. (eds.) ITS 2002. LNCS, vol. 2363, p. 158. Springer, Heidelberg (2002)
16. Jordan, P., Hall, B., Ringenberg, M., Cui, Y., Ros, C.: Tools for authoring a dialogue agent that participates in learning studies. In: Proceedings of Artificial Intelligence in Education (AIED), Los Angeles, July 2007, pp. 43–50 (2007)
17. Smith, V.L., Clark, H.H.: On the course of answering questions. Journal of Memory and Language (1993)
18. Brennan, S.E., Williams, M.: The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. Journal of Memory and Language (1995)
19. Dijkstra, C., Krahmer, E., Swerts, M.: Manipulating uncertainty: The contribution of different audiovisual prosodic cues to the perception of confidence. In: Proc. Speech Prosody (2006)
20. Saadawi, G.M.E., Azevedo, R., Castine, M., Payne, V., Medvedeva, O., Tseytlin, E., Legowski, E., azen Jukic, D., Crowley, R.S.: Factors affecting feeling-of-knowing in a medical intelligent tutoring system: the role of immediate feedback as a metacognitive scaffold. Adv. in Helth Sci. Educ. (2009)
21. Litman, D., Moore, J., Dzikovska, M., Farrow, E.: Using natural language processing to analyze tutorial dialogue corpora across domains and modalities. In: Proc. Intl. Conf. on Artificial Intelligence in Education (2009)
22. D'Mello, S.K., Craig, S.D., Witherspoon, A., McDaniel, B., Graesser, A.: Automatic detection of learner's affect from conversational cues. User Modeling and User-Adapted Interaction: Journal of Personalization Research 18, 45–80 (2008)