

Training of summarisation skills via the use of content based feedback

W. Lenhard, H. Baier, J. Hoffmann, W. Schneider, A. Lenhard

Abstract—conText is an intelligent tutoring system in German language that has been developed to improve text comprehension and writing abilities. The system is based on the model of Summary Street [4] and features comparable feedback mechanisms. Two laboratory experiments demonstrate that conText improves the quality of text summaries written by undergraduate students. We will assess the applicability of conText in schools in the following three years.

I. INTRODUCTION

Training of the ability to summarize texts is one of the most effective starting points for interventions aimed at fostering text comprehension [1]. Summarization not only results in a more active reading and a deeper level of processing. It also leads to a more integrated knowledge representation of a given text [2]. However, mere practise without supportive feedback is unlikely to produce a noticeable learning effect. Unfortunately, frequent and individualized feedback, albeit necessary, is hardly realisable under real life school conditions. In order not to pose additional work load on the teacher and to free resources for at-risk students, LSA-based tutoring systems may be used to assist the learning process.

In prior experiments in German language, LSA showed an equal performance compared to university students in classifying animal species and achieved medium to high correlations with human raters in essay scoring. Using added scores, the correlations reached values comparable to the reliability of standardized tests [3]. The correlation of the LSA-based scores with expert ratings varied between .6 and .8. Inter-rater correlation among expert was not significantly higher compared to the correlation between LSA and the experts. LSA therefore seems to be a valuable tool in the construction of intelligent tutoring systems not only in English, but also in German.

Following the work of Steinhart [2], [4] and the commercial application Summary Street (Pearson Knowledge

Technologies), an intelligent tutoring system named “conText” is currently under development for the German language [5]. It is aimed for the application in school grade 5 to 8 and features feedback on orthography, plagiarised passages, redundant and irrelevant sentences, as well as detailed and overall content coverage. Moreover, and contrary to Summary Street, it is planned to include feedback on style and composition as well and incorporates an educational text selection module that chooses the adequate degree of text difficulty on the basis of the student’s performance in preceding passes.

First experiments have been conducted to examine the effects of LSA-based feedback on the performance of students in writing summaries.

II. METHODS

We performed two experiments in which undergraduate students summarized an expository text. In experiments 1, there were two experimental groups: one that received LSA-generated content feedback, and one group that did not receive LSA-generated content feedback. In experiment 2, a third group that also received sentence analysis (feedback on potentially redundant and irrelevant sentences) was additionally examined.

A. Participants

In the first experiment, 20 students (10 male and 10 female) participated in the experiment. In experiment two, 52 students (12 male, 40 female) took part. The students were randomly assigned to the different experimental conditions and received course credit for their participation.

B. Apparatus and procedure

After receiving an instruction, the participants worked with a laboratory version of conText. The program first shows the source text, and a text processor component is displayed next. The students receive online-feedback on the length of the summary, the spelling and plagiarised passages. After a length-threshold has been met, an analysis on sentences is carried out where irrelevant and redundant sentences are flagged. Finally, the content coverage of the different passages of the source text, as well as a global rating are computed via LSA and displayed as vertical bar charts. The student may now stop or engage in another trial to further improve the draft. In experiment one, students worked on a text about earth quakes. In the second experiment, a text about moa (extinct flightless birds from New Zealand) was presented. Students had one

Manuscript received March 6, 2007. This work was supported by the German Research Foundation (DFG) under Grant HO 1301/11-2 and SCHN 315/29-1.

W. L. H. B., J. H., W. S. and A. L. are with the Department of Psychology, University of Würzburg, Röntgenring 10, 97070 Würzburg, Germany (W. L. phone: +49 931 312626; fax: +49 931 312763; e-mail: wolfgang.lenhard@mail.uni-wuerzburg.de; baier@psychologie.uni-wuerzburg.de; hoffmann@psychologie.uni-wuerzburg.de; schneider@psychologie.uni-wuerzburg.de; lenhard@psychologie.uni-wuerzburg.de).

hour time for the summarisation, but could stop working whenever they wanted to do so.

C. Measures

After completion of the experiments, participants rated the perceived accuracy of the content feedback of conText. The summaries were then scored by independent raters with a predefined rating scheme. Furthermore, time on task and the number of trials were recorded.

D. LSA-Platform

The LSA-platform consists of a server written in Java which runs on a desktop computer (Pentium IV, 3.2 GHz, 3 GB RAM, SuSE Linux 9.3). The server is administrated via a web interface [3] and deals with all aspects concerning the corpora administration, generation and weighting of frequency matrices, singular values decomposition (SVD), generation of semantic spaces and calculation of text similarities. The abridged Lanczos algorithm [6] is used for decomposing the singular values.

There are several client applications communicating with the server via internet, for example a system for automatic essay scoring of student writings in university lectures and laboratory prototypes of conText.

We use specialised semantic spaces for the computation of text similarities. The space underlying the described experiments consists of texts from the domains biology, geography and geology from school books, encyclopaedias and internet pages. The texts were extracted and split into paragraphs automatically. We converted all words in the texts to lower case and filtered stop words, words occurring less than three times as well as texts consisting of less than ten different words. The frequency matrix included 37 773 paragraphs with 83 369 different words (total size of corpus 2 178 432 words). Prior to the SVD, a log-entropy weighting was applied to the frequency matrix. We extracted 400 dimensions (duration of computation: 35min 17 sec) [3]. Other semantic spaces for different knowledge domains are at hand as well. When doing similarity judgments, missing words are automatically lemmatized in case, the semantic space contains the lemma.

III. RESULTS

Whereas all data assessed in experiment one could be used for analysis, a total of five participants were excluded from experiment two, either due to the lack of German language skills (N=3), or because technical errors occurred during the experiment (N=2).

The LSA-scores of the summaries significantly correlated with the human ratings, with $r=.552$, $p<.05$ in experiment 1, and $r=.738$, $p<.001$ in experiment 2.

In experiment 1, students who received LSA-feedback tended to receive better scores by human raters compared to students who did not receive feedback, $F(1, 18)=2.531$, $p=.064$. The effect size was $d(\text{Cohen})=.71$. Comparing the quality of their first and their final draft, they showed a higher

increase of content coverage in the course of writing, $F(1, 18)=4.514$, $p<.05$. This effect amounted to $d(\text{Cohen})=.95$.

In experiment two, due to a lower quality of the first draft, we failed to obtain a better final content quality in the groups who worked under LSA-based feedback as compared to the control group. However, the experimental groups showed a higher gain of content quality during writing than the control group, $F(2, 42)=4.34$, $p<.01$, with an effect size of $d(\text{Cohen})=.98$. The group receiving both content feedback as well as feedback on redundant and irrelevant sentences showed the highest increase in content coverage. The difference to the content feedback only group was not significant, however. When feedback was given, students worked longer, $F(2, 42)=3.71$, $p<.05$, and did more revisions, $F(2, 42)=64.02$, $p<.001$. Moreover, the participants' ratings of the feedback quality given by conText increased with detailedness, $\chi^2(2)=4.60$, $p=.05$. This finding underlines the usefulness of the sentence analysis in terms of user acceptance of the learning environment.

IV. CONCLUSION

LSA-generated feedback scores showed medium to high correlations with human judgments. Students, who work under LSA-feedback tend to write better summaries and show higher increases in the quality of their summary during writing. The obtained effects are encouraging, especially when keeping in mind that feedback in well-designed studies on average yields an effect size of .46 [7].

The next step will be to assess the effects in schools with first experiments in the summer of 2007, and to create a robust learning environment. In order to measure the long-term effects on reading comprehension, longitudinal studies will take place between autumn 2007 and 2010.

REFERENCES

- [1] E. Souvignier and F. Antoniou, "Förderung des Leseverständnisses bei Schülerinnen und Schülern mit Lernschwierigkeiten - eine Metaanalyse," *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, vol. 1, pp. 46-62, January 2007.
- [2] D. Wade-Stein and E. Kintsch, "Summary Street: Interactive computer support for writing," *Cognition and Instruction*, 22, 333-362, 2004.
- [3] W. Lenhard, H. Baier, W. Schneider and J. Hoffmann, "Automatische Bewertung offener Antworten," *Diagnostica*, in press.
- [4] D. Steinhart, "Summary Street: an Intelligent Tutoring System for Improving Student Writing through the Use of Latent Semantic Analysis," Ph.D. dissertation, Institut for Cognitive Science, University of Colorado, Boulder, CO, 2001.
- [5] W. Lenhard, H. Baier, W. Schneider and J. Hoffmann. (2006). conText: Fostering text comprehension by working with texts [Online]. Available: <http://www.summa.psychologie.uni-wuerzburg.de>
- [6] C. Lanczos, "An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators," [Online]. Available: <http://www.cs.duke.edu/courses/fall06/cps258/references/Krylov-space/Lanczos-original.pdf>, date of retrieval: 10.01.2007), *Journal of Research of the National Bureau of Standards*, 48, 255-282, 1950.
- [7] R. L. Bangert-Drowns, C. L. Kulik, J. A. Kulik and M Morgan, „The instructional Effect of Feedback in Test-Like Events,“ *Review of Educational Research*, vol. 61, 213-238, 1991.