Select-a-Kibitzer: A computer tool that gives meaningful feedback on student compositions

Peter Wiemer-Hastings
Arthur C. Graesser
The University of Memphis
Department of Psychology
Memphis TN 38152-6400
(pwmrhstn, a-graesser)@memphis.edu

September 9, 1999

Send correspondence to:
Peter Wiemer-Hastings
Department of Psychology
University of Memphis
Campus Box 526400
Memphis TN 38152-6400
(901)678-3529
(901)678-2579 FAX
pwmrhstn@memphis.edu

Submission for Special Issue of Interactive Learning Environments, guest edited by Joseph Psotka

running head: Select-a-Kibitzer

word count: 7000

Abstract

Select-a-Kibitzer is a computerized tool that gives feedback to students on their compositions in a unique way. The feedback is based on composition research which describes the process of writing as one of simultaneously solving multiple, possibly conflicting constraints. In Select-a-Kibitzer, each constraint is personified by a different character. A student enters a composition into the tool and then asks for feedback. A variety of natural language processing techniques are used to analyze the text. Then, each of the characters gives feedback on the text from its particular point of view. Select-a-Kibitzer differs greatly from standard "style checker" mechanisms that focus on surface features of the text. By using Latent Semantic Analysis, Select-a-Kibitzer can address a wide-range of meaning-oriented composition issues, including coherence, purpose, topic, and overall quality. This paper describes the composition research that forms the basis of the project, and the interaction and implementation of Select-a-Kibitzer. It focuses on techniques for using LSA to provide feedback about the meaning of the composition.

keywords: computer-aided education, composition, latent semantic analysis, natural language processing

1 Introduction

Writing is at the same time one of the easiest things that people do, and one of the most difficult. Children naturally move from talking about their knowledge to putting it on paper. But children's early texts lack the structure, coherence, and interestingness of mature texts.

Feedback on texts is necessary to help the student's writing processes mature, but good feedback is difficult to give. Because teachers are under enormous time pressures, they are faced with a choice: either assign less writing, or limit the time that they spend on evaluating student texts by giving quick feedback. Unfortunately, the easiest feedback to give is the most removed from the essential meaning of the texts.

Computers tools can help because they are relatively inexpensive and less time-constrained than teachers. But standard computer style-checking tools leave much to be desired. They tell when a word is spelled unlike any other word, and when a sentence is too long or in the passive voice. This type of surface-level, structural feedback is unlikely to significantly improve a student's writing. There are many aspects of the writing process that this approach leaves untouched. And because the main function of writing is to communicate meaning and/or feeling, feedback on semantic aspects of the writing is especially important.

As other papers in this issue demonstrate, Latent Semantic Analysis (LSA) (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Landauer & Dumais, 1997) enables a computer to make robust analyses of the meaning of a text. This paper describes a computer tool that uses LSA and other language processing techniques to give feedback on student compositions. It is based on Flower's view of composition as one of "negotiated construction of meaning" (Flower, 1994). This view stipulates that the process of composition requires the simultaneous solution of a variety of constraints. The constraints come from the writing task, the social context, the writer's view of the reader, and from the writer's knowledge state. These constraints are often contradictory, pushing the writer in different ways. The writer's task is then to make the necessary tradeoffs to find the solution that best reaches his

or her goals. The computer tool makes these constraints salient to the student by associating each one with a different character. The tool is unofficially called Select-a-Kibitzer (hereafter SaK).

Definition 1 *kibitzer: (n., informal, from Yiddish) 1. a spectator at a card game who looks at the players' cards over their shoulders, esp. one who gives unsolicited advice. 2. a giver of uninvited or unwanted advice. 3. a person who jokes, chitchats, or makes wisecracks, esp. while others are trying to work or to discuss something seriously* (Random House Webster's Unabridged Dictionary, 1997)

The basic premise of SaK is that if students receive a variety of different types of feedback on their compositions, and each type of feedback is associated with a different character (or agent), then the writers will better learn about the forces which make composition difficult, and will better learn how to make the necessary tradeoffs.

This paper initially describes prior research that forms the foundation of SaK, research into both the processes of composition and other computer mechanisms for supporting composition. Section 3 describes the high-level interaction between the student and the system. In section 4, we concentrate on the various ways in which LSA is used to give feedback on student texts, and in section 5, we describe the other technical underpinnings of the system. Finally, we discuss the status of the project and future research directions.

2 Background

As in the process of composition, there have been a variety of forces from different directions that have affected the development of SaK. This section describes the composition research that forms the theoretical foundation of the project. Then it describes related computer tools that give feedback on text and use agents in education.

2.1 Composition

Nearly two decades ago, Flower and Hayes shifted attention from the products of writing to the processes of writing (Flower & Hayes, 1981). Recently, Flower delved deeper into the nature of those processes (Flower, 1994). Her work, informed by Bakhtin (1981) and Nystrand (1986), characterizes writing as the negotiated construction of meaning. She describes a variety of voices that speak during the process of composition. These are the voices of the different types of constraints that impinge on the process. The term voice emphasizes that the constraints are not innocent bystanders to the process. Instead, they are actively involved, pushing the writer in different directions. SaK takes this notion a step further, giving each voice a face and a personality.

The negotiation aspect of the process comes in in two ways. First, the writer must negotiate with him or herself about the tradeoffs which will be necessary to best meet the constraints. Second, because writing is about the communication of meaning, the writer must negotiate with the reader about their common ground (Clark, 1996) and how the writer can attempt to pass on her knowledge or feelings.

Where do these voices come from? Some of them come from the reader. Some of them come from the particular situation in which the writing is carried out. Some of them come from the socio-linguistic context in which the student is writing. The biggest problem with these voices is that the writer is not normally consciously aware of them. She can't just write something that "sounds good." SaK's goal is to help the writer convert these ethereal forces into concrete considerations that she is consciously aware of and can then reason about. In Table 1, we list some of the implicit influences identified by Flower along with explicit questions that the writer can use to address them.

Beck and colleagues (Beck, McKeown, & Worthy, 1995) also use the voice metaphor to describe composition, but from the reader's point of view. Good writing speaks to the reader. This type of voicing isn't in the spelling of the words or the passive/active distinction. It's

Table 1: External influences and their related internal considerations

Source	External Influences	Internal Consideration
Reader	R's goal	What do I want the reader to know?
	R's knowledge	What do I know about what the reader knows?
	R's emotions and beliefs	What do I want the reader to feel about the subject?
	R's intentions	What do I want from the reader (esp. the teacher)?
Situation	Assignment	What am I supposed to write about?
	W's knowledge	What do I know about my topic?
	W's metaknowledge	What do I know about what I know about my topic?
	W's motivation	What do I want to say about the topic?
Socio-linguistic	W/R relationship	How do the social and linguistic relationships between writer and teacher/reader and writer and peers affect what will be said?
	Discourse	How do language/discourse conventions constrain affect what will be said?
	Language	What does my language allow me to express?

(R = Reader, W = Writer)

in the meaning. Beck *et al* take the notion of voice in part from Bakhtin. As they interpret Bakhtin,

> voice is "the speaking consciousness," in which the speaker is seen as a link in a chain of communication. There is no singularity of voice, but rather voice involves the simultaneous presence of social, cultural and political influences that have contributed to the speaker's perspective and world view (Beck et al., 1995, p. 224).

To take the voice metaphor to absurd lengths, SaK wishes to speak for the latent voices of the text. Students have a hard time separating their compositions from their own knowledge of what they wanted to say. If the SaK agents can tell the writer what they understand the text to be saying, then the writer can better perceive erroneous assumptions or poorly executed goals.

2.2 Agents in education

From an artificial intelligence point of view, the obvious way to implement Flower's "voices" is as agents. Each agent can embody or personify a particular voice. In other words, an agent can give a face, behavior, and personality to a voice, and thus make the different attributes associated with the agents more distinct and more memorable. This section describes other research on the use of agents in education.

Rickel and Johnson (1999) developed the STEVE (Soar Training Expert for Virtual Environments) model which teaches a student how to operate and maintain complicated machinery by actually demonstrating it in a three-dimensional virtual world. Using virtual reality technology, the student enters this world as well and can perform manual operations in the world. Thus, this environment allows a wide-range of instructional modes, including demonstration (with a talking head for speech output), "hands-on" training, and one-on-one feedback — the agent corrects the student if he/she makes a mistake and ensures that the equipment is configured to continue operation.

Lester and colleagues have developed a life-like animated agent, Cosmo, who inhabits a learning environment for the domain of Internet packet routing (Lester, Voerman, Towns, & Callaway, 1997). Their research focuses on using agents with "deictic believability" to pursue pedagogical goals. Cosmo's planning system selects and coordinates locomotive, gestural, and speech behaviors while explaining concepts like internet packet routing to students.

In meta-agent research, Paiva and colleagues are examining how to design agents that can

be used in many different educational tasks (Paiva & Machado, 1998; Paiva, Machado, & Martins, 1999). This project includes agents that perform different types of tasks, including learning about the user and acting as a pedagogically appropriate domain expert. The key to the approach is a protocol for communicating between agents so that they can cooperate effectively.

The Teachable Agents project turns the tables by asking high school students to teach math and science concepts to a computer agent (Brophy, Biswas, Katzlberger, Bransford, & Schwartz, 1999). Because the affable but naive agent will fail an "exam" if poorly taught, students are motivated to help. The system kindly offers resources that the students can explore to increase their own knowledge. When the students pass the knowledge on to the agent, their own understanding of the material is strengthened.

2.3 Agents and writing

In their "Little Planet" project, the members of the Cognition and Technology Group at Vanderbilt University (Cognition and Technology Group at Vanderbilt, 1996, 1998) have attempted to create an authentic task that will motivate young students (in the kindergarten or 1st grade level) to produce a "book" to achieve a communicative goal. Although the students do not actually write their own text for this book, they choose from sets of scenes, canned texts, and even music to tell a story about the scientific method so that others on the Little Planet won't be swindled. In this case, the characters in the simulated environment do not serve as teachers of any particular aspect of writing, but as motivation for the students to create an effective product.

2.4 Other computer support for composition

Kieras and Dechert (1985) surveyed pychological research on reading and extracted a set of comprehensibility principles. For example, although stylistic conventions suggest that

frequent references to the same item should be varied to make the text more interesting, psychological studies show that inconsistent reference decreases the readability of text. Based on this set of principles, Kieras developed a tool to improve the readability of technical documents for the U.S. Navy (Kieras, 1989). Kieras' tool included a special-purpose parser and a rule-based decision making system. It provided detailed feedback on the structural problems that it encountered in the text along with general suggestions for fixing them. It did not, however, comment on semantic aspects of the text.

Smith and Lansman (1989) developed a computerized tool that was intended just as much as a research tool for exploring composition process as as a tool for aiding those processing. There work was based on Hayes and Flower's theory of three major processes in composition: planning, rewriting, and revising. They focused on technical papers which are heavily reliant on a sound logical structure. Their computer tool provided graphical tools for brainstorming and structuring ideas in the planning process and for focusing on the structure as they turned it into text. This tool aided writing research by recording every step that the writers took. That gave the researchers a detailed view of the operations carried out in each different phase of writing.

Scardamalia and Bereiter's research has had a somewhat deeper focus, looking at collaborative "knowledge building" which uses written dialogue as the medium (Scardamalia & Bereiter, 1992, 1996). They have developed a system called CSILE (computer-supported intentional learning environments) which serves as a virtual forum in which students ask for information (from other students), offer their hypotheses, and "publish" their results with the help of teachers and classmates. This environment helps them learn not just the source material, but also strategies for learning.

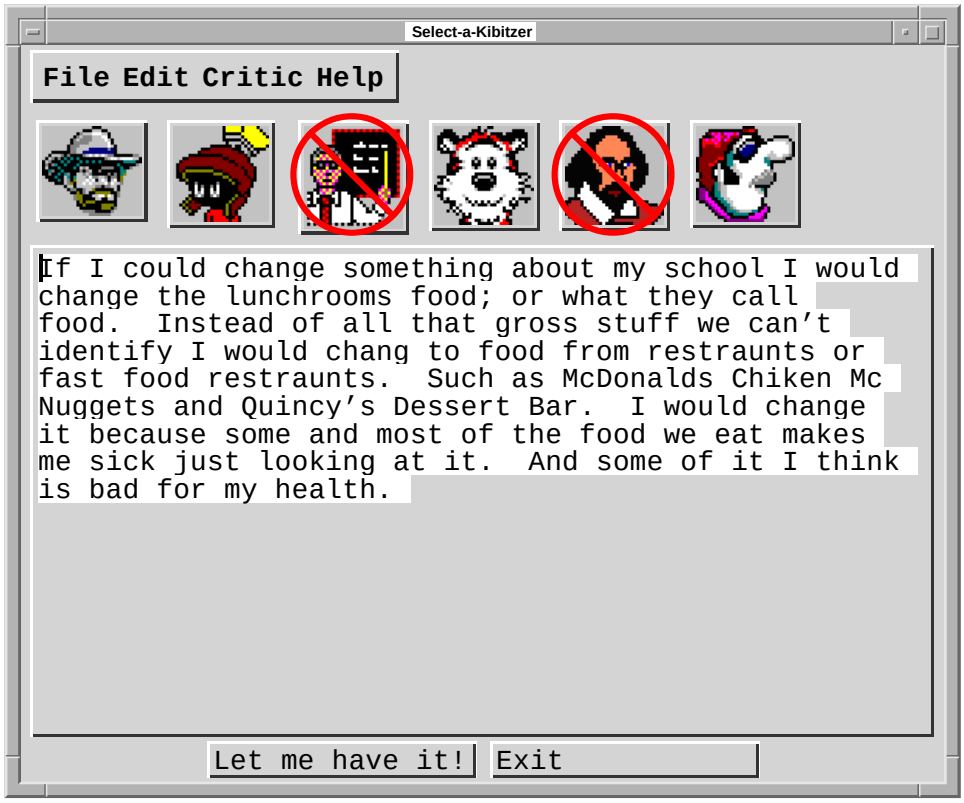

Figure 1: SaK main window

3 Interface and interaction

To (finally) give the reader a better idea about the nature of SaK, this section describes a student's eye view of the system. We include here some examples of the main SaK windows, and descriptions of the different ways in which students interact with the system. The details of the knowledge sources and processes which produce the feedback are given in the following two sections.

Figure 1 shows the main SaK window (including an authentic student text). The student enters her text into the central part of the window, which provides standard editing options. Just above the text are the pictures of the critics who are available to comment on the text. Clicking on a critic's picture toggles its activity status. Thus, the student can choose which aspects of the composition she would like feedback on. The menu at the top of the screen

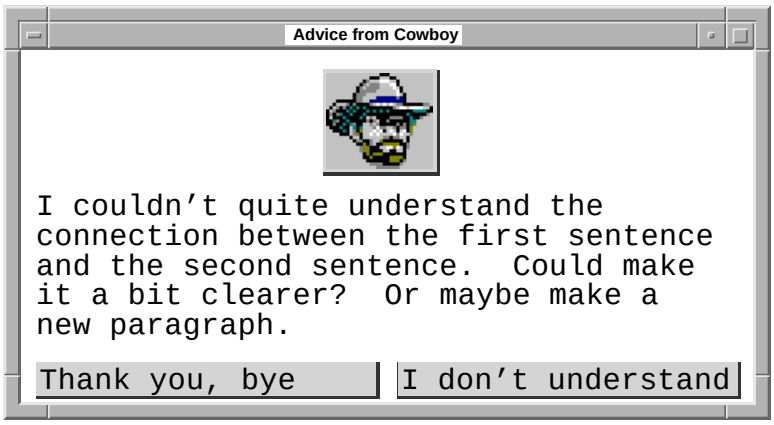

Figure 2: Advice from the cowboy

provides functions for loading and saving the text, performing various editing operations, and also help screens which describe the functioning of the system and the "personalities" of the individual critics.[1]

From the student's point of view, she presents her composition to one or more of a set of critics, and asks for feedback. Each critic in turn gives its feedback to the student. The purpose of the feedback is just to point out what is good or bad about the text, not to correct it. After getting the feedback, the student can revise the text and ask for feedback again. More specifically, the student first enters the composition into the system, either by typing or by cutting and pasting. Of course, this requires that the students have keyboarding abilities, but this seems to be becoming more common in students. Speech understanding technology will be added to the system when the technology becomes sufficiently mature (Karat, Halverson, Horn, & Karat, 1999, for a pessimistic prognosis).

Figure 2 gives an example of one of the feedback windows. The cowboy is interested in the coherence of the text. He uses LSA (as described in Section 4.2), to identify places in the text which don't seem to fit together. The feedback window gives a brief explanation of what the critic sees as the difficulty and also specifies its location. The student can then dismiss

[1]In the current system, we did not make a serious attempt to associate the different critics with stereotypically "appropriate" personalities. It could be that doing so would help students determine what the underlying concept is, but we do not want to resort to using stereotypes unless absolutely necessary.

Figure 3: A "push" window

the window or ask for further explanation about that topic.

The personality of the kibitzers is exhibited only by the type of feedback that they make and the picture on this window. Ideally, this association between character and text aspect would be augmented by other agent attributes. The inclusion of a distinctive voice for each agent and perhaps some animation could make the agents much more distinguishable. These feature will be implemented in a future version of the system.

Another type of interaction that is supported is the "push" (Graves, 1983). For certain types of problems, students can rewrite parts of their compositions "offline". Figure 3 shows feedback from Marvin which presents the student with a portion of her original text and asks her to modify it. She can then edit the original text, make one or more completely new versions, or maintain the original. Then she can have the new text spliced back in to the full composition.

In summary, the student's interaction with the tool is fairly simple. The student enters the composition and asks for advice. The active critics offer their feedback, giving further explanation when necessary. The complexity of the system lies behind the scenes, as is

described in the next sections.

4 Semantic feedback using LSA

A crucial goal of this project is to give the feedback on aspects of writing that is relatively difficult to formulate and is very important to the communicate goals of the text: feedback on the meaning of the composition. In this section, we focus on the various ways that LSA allows SaK to perform semantic evaluations of texts.

4.1 LSA background

Because the technical details of LSA have been described elsewhere (Deerwester et al., 1990; Landauer & Dumais, 1997, for example), and in other articles in this issue, we will not go into the details here. We do want to emphasize some important points. What LSA does is to give measurements of the semantic similarity between texts. Although at first glance, this seems rather limited, there are a surprising number of text evaluation tasks which can be framed in terms of semantic similarity. Once the task is appropriately defined, the researcher's task is to find or create appropriate texts to serve as the corpus and comparison texts for LSA. The following sections provide examples of such endeavors.

We have previously used LSA to evaluate student contributions in the context of an intelligent tutoring system (Graesser, Franklin, Wiemer-Hastings, & the Tutoring Research Group, 1998; Wiemer-Hastings, Graesser, Harter, & the Tutoring Research Group, 1998). In addition to demonstrating that LSA can rate the quality of student texts as reliably as intermediate-knowledge human raters can, we examined a variety of aspects that affect its performance. We studied the effects of the size and composition of the training corpus (Wiemer-Hastings, Wiemer-Hastings, & Graesser, 1999b), the sensitivity of LSA to various features of the student texts (such as the presence of negation, which has no significant

effect), and compared its performance to a version of LSA without SVD and to a keyword-matching mechanism (Wiemer-Hastings, Wiemer-Hastings, & Graesser, 1999a). Relevant details of those analyses are given in Section 4.7. The work described here was influenced by this tutoring work, and also by the work of Foltz (Foltz, 1996; Foltz, Kintsch, & Landauer, 1998, and this issue) and Kintsch (this issue) on evaluating student essays. The rest of this section describes the various functional roles that LSA plays in SaK. As example texts, we will use samples of writing from 6th graders who were given the following assignment as part of state-wide testing:

> If you could change something about school, what would you change? Write a few paragraphs that answer the question. In your answer be sure to include: What would you change; why would you change it, and how would you change it. Use examples and details in your writing.

4.2 Intersentence and whole-text coherence

One aspect of text quality that LSA is well-equipped to comment on is coherence. The technique is simple: calculate the cosine between each pair of adjacent sentences. The ideal value should neither be too high nor too low. A very high cosine would indicated excessive redundancy between the sentences. A low cosine indicates a conceptual shift in the text. The text is either not well connected, or the student has started a new topic. As shown in Figure 2 above, SaK gives feedback addressing both of these cases.

By taking the average cosine between adjacent sentences, SaK gets a measure of the overall coherence of the text. For such global measures, the feedback given by SaK asks the student to look over the entire text to look for ways to make it more coherent. If the student requests, SaK will give examples of more and less coherent texts.

4.3 Purpose of sentence

One of the biggest complaints of composition teachers is that students often throw in topics that have no apparent connection to the assigned subject. The example assignment given above is one where the purpose of the student's text is relatively constrained. In a task like this, LSA can be used to determine the purpose of each sentence by comparing them with templates such as,

- **What:** "I would change," or more specifically, "I would change the food," "I would change the teachers," "I would change the classrooms," etc
- **Why:** "I would change it because," or just "because"
- **How:** "I would change it by," "How I would change it"

These template phrases are obviously very similar to the text of the assignment above. Young students are quite likely to produce text which closely matches the template of the assignment. For example, of 67 student texts addressing the school change assignment, 24 of them (35%) start with a phrase very much like "If I could change something about school I would ...". It should be noted that some of these template sentences are very similar to others and will result in similar cosines. A sentence like, "I would change the food by getting it from McDonalds" has an LSA cosine of 0.68 with "I would change", and a cosine of 0.69 with "I would change it by".[2] Although these cosines are both above a reasonable threshold for matching, they can be evaluated relatively. Because the latter threshold is slightly higher, the purpose of addressing how the school should be changed should be assigned to this sentence.

[2]When training LSA, it can be told to ignore certain words. The standard LSA implementation comes with a file containing 440 very common words including each of the words in the template sentences above except for "teachers" and "classrooms". Obviously this feature must be turned off when using LSA to match such sentences.

4.4 General topic

In a directed assignment like the example above, the students are free to write about a wide range of topics. But the topics that they choose tend to fall into a rather small range of subjects. For example, in the 67 compositions mentioned above, 17 mention the food, 21 mention the teachers, 5 mention dress code restrictions, 5 mention school hours, and 6 mention the other students. To perform topic matching for a given topic, a simple qualitative analysis is performed to identify the main topics of compositions. Then, the most prototypical sentences of each topic are picked as the comparison texts. Alternatively, clustering methods can be used on the trained LSA space to automatically locate semantic clusters in the corpus. The best prototype for each cluster is easily found by choosing the text with the highest average cosine with the other texts in the cluster.

With these sets of topics and prototype sentences in hand, SaK can give feedback in a variety of ways. If the student's composition (sentence-wise or as a whole) doesn't match any of the prototypes above a threshold, then a critic will just say that the topic is unclear and ask for more explanation. If the student text does match one or more prototypes, then different types of feedback are available. For example, a critic can suggest a related topic: "I agree that the food is terrible. And what about the condition of the cafeterias?" The system can also ask the student if they would like to see other compositions on the same topic. If the teacher can specify a target set of topics for the assignment, this mechanism can be used to ensure that all of the topics are addressed or to suggest inclusion of those that aren't.

4.5 Overall quality

As is described by Foltz (this issue), an overall assessment of a composition can be calculated by comparing it to pre-graded texts. For SaK, this is not done primarily for the purpose of reducing the grading load of the teacher. Instead, it is used as a general indicator of text quality. If a composition receives a low overall rating, then a range of feedback is possible.

The simplest would be to suggest that the composition seems to need more work and that the student should spend more time on it. As mentioned above, a kibitzer can suggest that the student add a topic that was not found in the text but which is present in one or more high-quality compositions, for example, "What do you think about the number of students in your classes?"

4.6 Automatic summaries

As previously mentioned, Beck and colleagues pointed out the importance of voice in text (Beck et al., 1995). SaK can speak for the text in a slightly different way than Beck intended by using LSA to automatically produce a summary of the text.[3] The coherence mechanism described above can break down the text into discrete semantic chunks. Then, the central sentence of each chunk is identified by taking the sentence with the highest average to the other sentences within a chunk. A kibitzer can present these key sentences to the user as its understanding of the main points of the composition. This technique can help the student take a fresh look at the composition, focusing on how the reader would understand it. This is a very important aspect of composition for a student to learn. As Beck et al state, "Viewing discourse as a chain of communicative discourse exists only insofar as it comes from somewhere and is addressed to another, being produced in the anticipation of a response." (Beck et al., 1995, p. 224)

4.7 Issues on LSA and composition

The preceding sections have described a wide range of evaluations of compositions that can be accomplished with the use of LSA. They aren't especially novel techniques, but the combination of them in such a tool for giving feedback on the various aspects of composition is an exciting new technology for the future. The descriptions above give a general idea

[3]This technique was suggested by Gerry Stahl, personal communication, October, 1998.

of how the different techniques are implemented, but there are several issues related to the writing situation and the LSA implementation that affect LSA's performance. Because SaK could be ported to a variety of different writing tasks, we include a discussion of those issues here.

LSA is a corpus-based technique, so the most obvious factor that affects it is the training corpus. The best corpus is one that is specific enough to allow subtle semantic distinctions within the domain, but is general enough that moderate variations in terminology won't be lost. Following advice from Foltz (personal communication, October 1996), we have set up training corpora that contain texts from a range of sources within the domain of interest. In our tutoring task, we found that a mixture of 40% of the texts from the general domain (computer literacy) and 60% in the specific tutoring areas (hardware, software, and the internet) gave the best overall performance (Wiemer-Hastings et al., 1999b).

The size of the training corpus for LSA is also very important. In related evaluations, we found that LSA worked best with the largest training corpus that we tested which consisted of 2 textbooks and 30 articles, for a total of 2.3 MB of text. But LSA showed graceful degradation; the performance for a "minimal" corpus which contained only 15% of the full corpus was only about 12% lower than the maximum performance for the entire corpus (Wiemer-Hastings et al., 1999b).

It is not easy to compile a domain-specific corpus, so some researchers have explored the use of general-purpose corpora. For example, researchers at the University of Colorado have trained LSA on a number of different corpora collected by a private firm. These corpora approximate the exposure to print of 3rd, 6th, 9th, and 12th graders as well as college students. A web-based interface allows comparisons of texts within the different LSA spaces.

The number of dimensions in the LSA space is also important. Most LSA applications use a dimensionality between 100 and 400, but the best number of dimensions must be empirically determined with respect to a particular domain and task. The ideal dimensionality is related to the size of the domain, both in the amount of text and the number of different terms in

the corpus. It is also related to the granularity of knowledge discrimination desired. For example, a corpus developed for a college computer literacy course would not be appropriate for analyzing a 4th grader's text about computers. The corpus would contain information about many concepts that the younger student would be unlikely to mention, and would therefore not have sufficient discriminative power for more relevant topics.

There are also a number of task-specific considerations that can affect the way that LSA is trained and used. The breadth of a given assignment affects what type of training corpus will be effective. If students must write about a tightly constrained topic (as in the school change example), a special-purpose corpus can be compiled from related texts and from sample compositions. And because it contains just the types of texts that the students should write about, it will be more effective at performing the similarity comparisons. If the writing task is not narrowly defined, a larger, less specific corpus is necessary.

The structural aspects of the writing assignment are also important. In our example task, the students were told to include information about what they would change, why they would change it, and how. SaK can use this information to provide feedback specific to the task. Some of the techniques described above may not be relevant for less strict writing tasks, and other tasks may lend themselves to additional types of feedback.

Finally, the education level of the student must be taken into account. The age of the student will affect the desired wording of the feedback, the types of feedback given, and the characters that the system uses. These considerations must be taken into account to provide a pedagogically valuable tool.

5 Other language analysis mechanisms

In addition to LSA, SaK uses a variety of other linguistic resources and language processing techniques to gain information on which it can base its feedback. This section describes those other inputs to the feedback process.

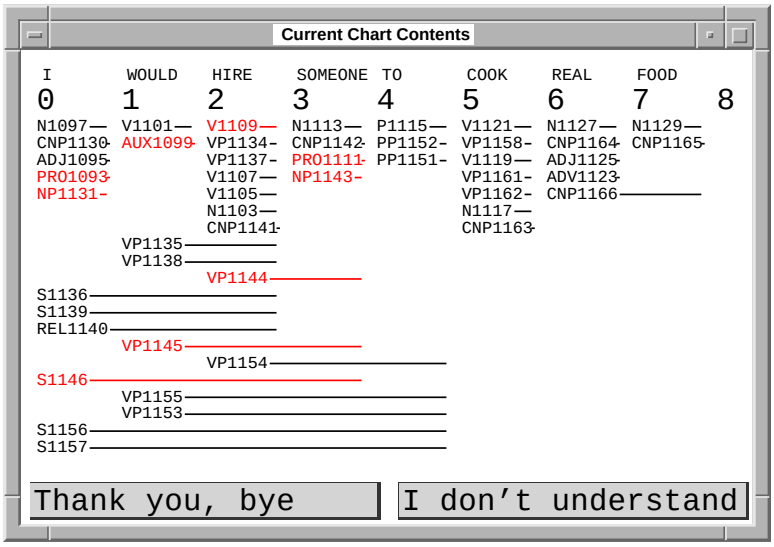

Figure 4: Display of the parser chart

5.1 Allen's Parser

For the syntactic analysis module of SaK, we have chosen a parser that is freely available and is associated with a standard Natural Language Processing textbook (Allen, 1995). The parser is a standard chart parser with a simple grammar formalism. The parser comes with a small set of grammar rules that can handle basic linguistic structures. We chose a standard parsing mechanism so that the grammar can be extended by future research projects. The lexicon and semantics of the system are described in the next section.

The syntactic results of the parser are used to produce feedback on grammatical well-formedness of the student's text. If the parser can not create a full parse for a sentence, then a kibitzer suggests that it is ungrammatical. Figure 4 shows the contents of the parser's chart for a sample sentence. The words in the sentence are displayed across the top of the window, and the arcs in the chart are shown underneath, with labels for the type of constituent, and lines to show the extent of the arc. A full parse is represented by an S (for sentence) link that spans the entire sentence. There is no such arc in Figure 4. Even though this sentence is grammatical, it is beyond the scope of the current, relatively simple grammar for the parser. Thus, it is especially important to have the semantic evaluations given by

LSA, which ignores grammar altogether.

5.2 WordNet as online lexicon

In order to avoid painstaking knowledge engineering of a hand-built lexicon for the system, we have created an interface between the parser and the WordNet system (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990; Fellbaum, 1998). WordNet was developed as a lexicographer's tool and provides information about four types of words: nouns, verbs, adjectives and adverbs. Its lexical knowledge includes information about exceptional spellings of these words for use in morphological analyses.

WordNet was intended as a tool for lexical research. Because it was meant to be independent of any semantic theory, its semantic representation consists solely of a synonym set (or "synset") for each different sense of a word. This set is intended only to allow a human user to determine which sense of the word is being refered to. For use with LSA, this type of semantic representation also works well. The synsets can support semantic restrictions, using LSA to judge the acceptability of attachments by the use of similarity instead of subsumption in an ontology. Alternatively, the hypernyms (more general terms) that WordNet supplies can be used for this purpose.

WordNet also supplies information about the "familiarity" of a word, defined on a scale relative to how many different senses the word has. SaK uses this information to give feedback on words that might not be known to all readers.

The basic information provided by WordNet must be substantially manipulated and augmented to be of use to the parser. Our extensions provide information about other word classes, for example, prepositions and pronouns. Words which occur in WordNet are coerced into the form required by the Allen parser. In particular, different senses of words which have the same syntactic profiles are combined into a single sense which has a disjunctive semantic represention. This simplifies the syntactic processing by reducing (syntactic) am-

biguity. WordNet provides coarse-grained semantic templates for verbs. Although this has not yet been implemented, these could be used for basic semantic constraint-checking in the parser. LSA can also be used to reduce semantic ambiguity by pruning the set of synsets provided by WordNet (Hu & Graesser, 1998). We are examining this approach in related research.

We have embedded the WordNet database in an object-oriented lexicon server, which allows for different types of information from different sources, and is extensible to include information on domain specific, closed class, and other words that are not included in WN. Using this mechanism, a single query provides all information about the (possibly ambiguous) senses of the word.

5.3 Soar decision making

Lexical information, parse information, and LSA information is put into the working memory of the Soar system (Laird & Rosenbloom, 1996). This information is represented as features associated with a position in the text. Soar rules determine if the individual kibitzers are active and decide what feedback to give at any particular time. The basic decision making process is very simple: if working memory contains a feature that has been identified as problematic, then tell the relevant kibitzer to give feedback. More complex rules could be implemented to take into account a number of different things. For example, the teacher can specify that a particular student needs to focus on a particular aspect of composition. The Soar rules can also implement student modeling by taking into account the identity of the student. The system could keep track of what feedback it has already given to a student (in this or previous sessions) to enable it to modify the feedback over time ("Like I told you before, ...").

6 Current status and future work

SaK is a system which uses a variety of powerful text analysis mechanisms, including LSA, to give feedback on student compositions. Each different type of feedback is presented to the student from a different character. Associating the characters with the aspects of composition should allow the students to better understand the competing forces that are involved in the process of writing. And because the kibitzers have different opinions and may even disagree with each other, the students should come to realize that the constraints are not hard constraints, but allow and require the student to make tradeoffs.

All of the technical aspects of the system are currently in place, but some pedagogical issues remain to be addressed. The following questions are beyond our current knowledge, but we will address them in future collaboration with experts in composition research:

- How much feedback can the students handle at one time?
- What is the best way to describe the various types of problems to students of different age groups?
- What types of feedback are appropriate for a given age group?
- What is the best way to achieve a balance between:
  - simple and complex feedback?
  - structural and semantic feedback?
- How can we best integrate teacher and curriculum goals into the system?

Once these questions can be addressed, we will be ready to assess the system's effects on students' composition processes. We assume that by promoting the association between the different characters and the different constraints on the composition process, we can increase the student's ability to recognize those constraints. We will assess this assumption and this

approach by comparing the compositions of students who use the tool, and student who perform a control task. If our approach is correct, the SaK students will be more aware of the different constraints that they are facing, they will be better able to take them into account when they are writing, and they will produce higher quality compositions.

Acknowledgments

This work was supported by a grant from the James S. McDonnell Foundation's Cognitive Studies for Educational Practice Program. Many thanks to invited kibitzer Katja Wiemer-Hastings for her helpful comments on many aspects of an earlier draft of this paper. Thanks to Linda Flower for helpful advice at the beginning of the project. And thanks to Susan Goldman, Joan Getschman, Anne Britt, and Clara Halamka for (anonymous) samples of student writing. The authors would also like to acknowledge the developers of the component technologies that we have imported into SaK. SaK runs on the Linux operating system, Franz Inc's free version of Allegro Common Lisp for Linux, and the Garnet user interface system from Carnegie Mellon University's Human Computer Interaction Institute. The LSA software has been made freely available by Bellcore for education purposes, and an initial Lisp version of the LSA comparison software was provided by Gerry Stahl. WordNet is available from Princeton. Allen's parser is available from the publisher of his textbook. The version of Soar used here is a retro version of modern Soar implemented in Lisp by Jans Aasman.

References

Allen, J. (1995). *Natural Language Understanding.* Addison-Wesley, Reading, Mass.

Bakhtin, M. (1981). *The dialogic imagination.* University of Texas Press, Austin. C. Emerson and M. Holquist, Trans.

Beck, I., McKeown, M., & Worthy, J. (1995). Giving a text voice can improve students' understanding. *Reading Research Quarterly*, *50*(2), 220–238.

Brophy, S., Biswas, G., Katzlberger, T., Bransford, J., & Schwartz, D. (1999). Teachable Agents: Combining Insights from Learning Theory and Computer Science. In Lajoie, S., & Vivet, M. (Eds.), *Artificial Intelligence in Education*, pp. 21–28 Amsterdam. IOS Press.

Clark, H. H. (1996). *Using language*. Cambridge University Press, Cambridge.

Cognition and Technology Group at Vanderbilt (1996). A Multimedia Literacy Series. *Communications of the Association for Computing Machinery*, *39*(8), 106–109.

Cognition and Technology Group at Vanderbilt (1998). Designing environments to reveal, support, and expand our children's potentials. In Soraci, S. A., & McIlvane, W. J. (Eds.), *Perspectives on fundamental processes in intellectual functioning: A survey of research approaches*, Vol. 1, pp. 313–350. Ablex, Stamford, CT.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, *41*, 391–407.

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. MIT Press, Cambridge, MA.

Flower, L. (1994). *The construction of negotiated meaning: A social cognitive theory of writing*. Southern Illinois University Press, Carbondale, IL.

Flower, L., & Hayes, J. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*(1), 365–387.

Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual coherence with Latent Semantic Analysis. *Discourse Processes*, *25*, 285–308.

Foltz, P. (1996). Latent semantic analysis for text-based research. *Behavior Research Methods, Instruments, and Computers*, *28*, 197–202.

Graesser, A. C., Franklin, S. P., Wiemer-Hastings, P., & the Tutoring Research Group (1998). Simulating smooth tutorial dialogue with pedagogical value. In *Proceedings of the 11th International Florida Artificial Intelligence Research Symposium Conference*, pp. 163–167. AAAI Press.

Graves, D. (1983). *Writing: Teachers and children at work*. Heinemann, Portsmouth, NH.

Hu, X., & Graesser, A. (1998). Using WordNet and Latent Semantic Analysis to Evaluate the Conversational contributions of learners in tutorial dialogue. In *Proceedings of ICCE'98,*, Vol. 2, pp. 337–341 Beijing. China Higher Education Press.

Karat, C., Halverson, C., Horn, D., & Karat, J. (1999). Patterns of entry and correction in large vocabulary continuous speech recognition systems. In *Proceedings of the Computer Human Interaction Conference* Washington. Association for Computing Machinery.

Kieras, D., & Dechert, C. (1985). Rules for comprehensible technical prose: A survey of the psycholinguistic literature. Tech. rep., University of Michigan.

Kieras, D. (1989). An advanced computerized aid for the writing of comprehensible technical documents. In Britton, B., & Glynn, S. (Eds.), *Computer writing environments: Theory, Research, and Design*. Erlbaum, Hillsdale, NJ.

Laird, J., & Rosenbloom, P. (1996). The evolution of the Soar cognitive architecture. In Steier, D. M., & Mitchell, T. M. (Eds.), *Mind Matters: A tribute to Allen Newell*, pp. 1–50. Erlbaum, Mahwah, NJ.

Landauer, T., & Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.

Lester, J., Voerman, J., Towns, S., & Callaway, C. (1997). Cosmo: A Life-like Animated Pedagogical Agent with Deictic Believability. In *Working Notes of the IJCAI '97 Workshop on Animated Interface Agents: Making Them Intelligent or Special Issue of the Applied Artificial Intelligence Journal*, pp. 61–69 London. Taylor and Francis.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). Five papers on WordNet. Tech. rep., Cognitive Science Laboratory, Princeton University.

Nystrand, M. (1986). *The structure of written communication: Studies in reciprocity between writers and readers*. Academic Press, New York.

Paiva, A., & Machado, I. (1998). Vincent: An autonomous pedagogical agent for on-the-job training. In Goettl, B., Halff, H., Redfield, C., & Shute, V. (Eds.), *Intelligent Tutoring Systems, Proceedings of the 4th International Conference*, pp. 584–593 Berlin. Springer.

Paiva, A., Machado, I., & Martins, A. (1999). Using and re-using agents in multi-agent learning environments. In Lajoie, S., & Vivet, M. (Eds.), *Artificial Intelligence in Education*, pp. 750–752 Amsterdam. IOS Press.

Random House Webster's Unabridged Dictionary (1997). Random House, New York.

Rickel, J., & Johnson, W. (1999). Virtual Humans for Team Training in Virtual Reality. In *Proceedings of the Ninth International Conference on AI in Education*, pp. 578–585. IOS Press.

Scardamalia, M., & Bereiter, C. (1992). Text-based and knowledge-based questioning by children. *Cognition and Instruction*, *9*(3), 177–199.

Scardamalia, M., & Bereiter, C. (1996). Student communities for the advancement of knowledge. *Communications of the Association for Computing Machinery*, *39*(4), 36–37.

Smith, J., & Lansman, M. (1989). A cognitive basis for a computer writing environment. In Britton, B., & Glynn, S. (Eds.), *Computer writing environments: Theory, Research, and Design*. Erlbaum, Hillsdale, NJ.

Wiemer-Hastings, P., Graesser, A., Harter, D., & the Tutoring Research Group (1998). The foundations and architecture of AutoTutor. In Goettl, B., Halff, H., Redfield, C., & Shute, V. (Eds.), *Intelligent Tutoring Systems, Proceedings of the 4th International Conference*, pp. 334–343 Berlin. Springer.

Wiemer-Hastings, P., Wiemer-Hastings, K., & Graesser, A. (1999a). How Latent is Latent Semantic Analysis?. In *Proceedings of the Sixteenth International Joint Congress on Artificial Intelligence*, pp. 932–937 San Francisco. Morgan Kaufmann.

Wiemer-Hastings, P., Wiemer-Hastings, K., & Graesser, A. (1999b). Improving an intelligent tutor's comprehension of students with Latent Semantic Analysis. In Lajoie, S., & Vivet, M. (Eds.), *Artificial Intelligence in Education*, pp. 535–542 Amsterdam. IOS Press.