Journal of Universal Computer Science, vol. 7, no. 3 (2001), 278-289
submitted: 27/12/00, accepted: 13/3/01, appeared: 28/3/01 *© Springer Pub. Co.*

An Online Writing Platform for Language Learners[1]

David Wible
(Graduate Institute of Western Languages and Literature
Tamkang University
Tamsui, Taipei Hsien, Taiwan
dwible@mail.tku.edu.tw)

Chin-Hwa Kuo
(Computer and Networks (CAN) Laboratory
Department of Computer Science and Information Engineering
Tamkang University
Tamsui, Taipei Hsien, Taiwan
chkuo@mail.tku.edu.tw)

Nai-lung Tsao
(Computer and Networks (CAN) Laboratory
Department of Computer Science and Information Engineering
Tamkang University
Tamsui, Taipei Hsien, Taiwan
889190129@s89.tku.edu.tw)

Anne Liu
(Graduate Institute of Western Languages and Literature
Tamkang University
Tamsui, Taipei Hsien. Taiwan
lieor@ms7.url.com.tw)

Abstract: The purpose of this paper is to describe one module in a highly integrated language learning environment. The module described is an asynchronous interactive online environment for EFL writing which integrates the potential of computers, Internet, and linguistic analysis to address the highly specific needs of second language composition classes. The system accommodates learners, teachers, and researchers. A crucial consequence of the interactive nature of this system is that users actually create information through their use, and this information enables the system to improve with use. In addition to the tools provided for teachers to mark essays and automatically track the feedback they have given each learner, the system supports the automated capture of a learner corpus of written English in the process. The essays written by users and the comments given by teachers are archived in a searchable online database. Learners can retrieve this information to examine their own recurring problems in the target language. Teachers can do the same in order to discover these problem areas for individual learners and for a class as a whole. The modular system provides interfaces with functions to facilitate an array of user tasks such as teachers' correction of

[1] This is an extended version of a paper presented at the ICCE/ICCAI 2000 conference in Taipei, Taiwan. The paper obtained an Outstanding Paper Award and is published with the permission of ICCE/ICCAI.

essays and learners' writing and revision processes. Error analysis of learner essays has led to content creation for automated online help. One sort of help feature can detect certain errors automatically and offer appropriate help pages. Another type of help feature can track the number of times a teacher has marked the same error type in one learner's writing and, when this number reaches a threshold, automatically offer help on this error to this learner.

Keywords: EFL writing, CALL, corpus linguistics, web-based learning environments
Category: K.3

1 Introduction

The purpose of this paper is to describe one module in a highly integrated language learning environment called IWiLL (Intelligent Web-based Interactive Language Learning: http://iwillnow.org). The module within IWiLL which we focus on in this paper is a novel web-based writing environment designed for EFL composition classes.

The design of the IWiLL writing environment is intended to meet certain criteria. It supports interactivity between teacher and student and among students. In addition, the system takes advantage of the computers' capacity to track the content of the interactions between users and to enable users to do pinpoint searches of the record of these interactions. This feature offers invaluable information that can serve as a source of insight for both learners and teachers, information which in traditional writing classrooms remains out of reach since it is highly distributed.

While the system described here is designed for second language writing classes, it is more accurately seen as one component in an integrated language learning environment that includes other skills, such as reading and listening [Wible et al (00); Kuo et al (00a); Kuo et al (00b)]. The modularized and integrated design is intended to accommodate recent trends in language pedagogy which view language skills as best learned in an integrated whole rather than as a set of separate, independent skills. Finally, while the IWiLL environment is designed specifically to meet the needs of certain type of language course (second language composition) it is intended to provide as much flexibility as possible so that teachers are free to use their own approaches and the materials of their choice. In other words, the system offers a platform compatible with any variety of curricula or syllabi.

2 The Organization of the System

2.1 The Teacher's View

We consider the system first from the point of view of a teacher. A registered teacher who logs onto the system is presented a display screen of various links to components within the system. To correct student essays, the teacher links to a page which displays their student roster. Each essay that a student has turned in over the system is represented as a button appearing beside that student's name on the roster. From this roster screen, the teacher retrieves the essay by clicking on the button that represents that essay on the roster page. To mark the essay with a comment (for

example, to mark a run-on sentence or subject-verb agreement problem), the teacher first chooses the portion of the essay targeted for comment using the mouse. Once the relevant text has been selected, there are two ways for the teacher to provide the student with a comment on it. The first is to simply type the comment in the empty text box provided especially for the teacher's comment and then, once the comment has been composed, append it to the intended portion of the student's text by clicking on the appropriate button ('Give the comment'). The second way of providing a comment is to choose one that has been stored in a "Comment Bank." This second way deserves some elaboration.

The Comment Bank provides each teacher with a convenient means for storing and reusing frequently used comments. To retrieve a stored comment and append it to the portion of the student essay, the teacher simply selects that comment from a drop-down menu and clicks on it. The teacher can add new comments to her Comment Bank at any time [see Figure 1].

At this stage, research is needed to understand the factors effecting how beneficial various sorts of comments are in helping students with their writing. An advantage of this system is that, with it, researchers can control the crucial variables (such as the precise form and content of the teacher feedback being investigated), and it makes readily available the data needed for such research since the marked and unmarked essays are archived in forms that can be queried. Moreover, the revised versions of an essay can be examined along side the teacher's comments that were given to the student on the original version of the essay, making it possible to easily track the influence of various types of teacher feedback.

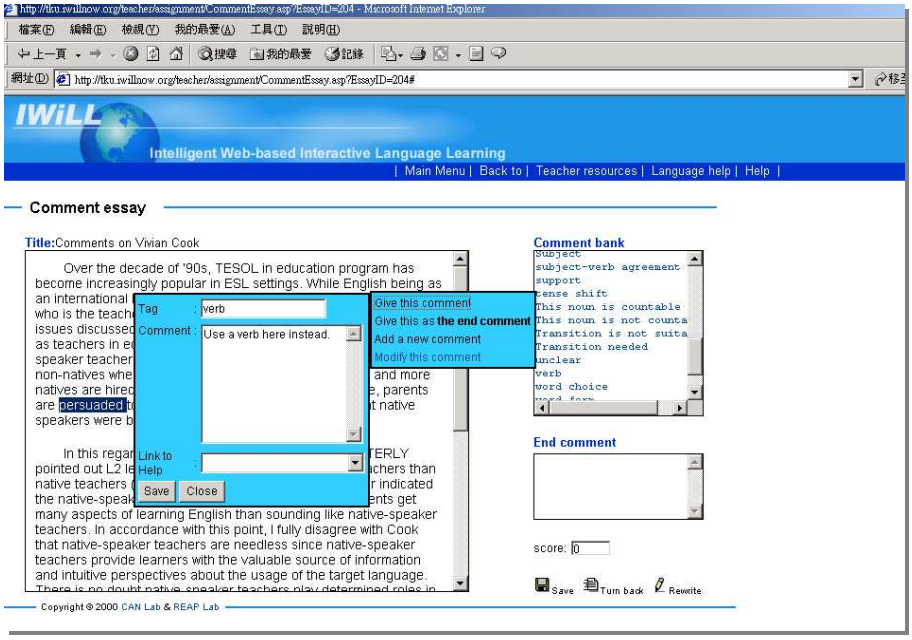

Figure 1: *Teachers' interface for marking student essays.*

It is important to notice the distinction between this essay-marking function and the superficially similar functions offered in commercial word processors such as Microsoft Word. Like our system, those programs allow users to select portions of text and annotate them with comments. While the convenience that this provides to users as a communication tool is essentially the heart of the function's role in these commercial word processing systems, in our system this convenience is a relatively incidental (though valuable) advantage. For us, the substantial value comes from a set of related features which the word processing programs do not offer. Specifically, all of the annotations provided in our system by teachers when they mark essays are permanently indexed, by way of database technology, to the portions of text that the teacher has marked. Moreover, the essays themselves along with the indexed teacher comments enter a permanent corpus of learner essays that can be searched on line. Information extraction techniques, then, make it possible to provide learners and teachers with instant cumulative profiles of the trouble spots of individual learners, of whole classes of learners, or subtypes of learners selected by a wide variety of criteria. For example, the system enables teachers to retrieve all tokens that have been marked with a particular error type either from the essays of a single learner or from the essays of groups of learners. Moreover, teachers can retrieve the tokens of every error type and display them in order of frequency, with the error type that has been marked on the highest number of text portions listed first.

The role which our commenting function can play is deepened greatly by the highly integrated nature of our system design. Not only does it support profiles of entire groups of learners, but the analysis of the common errors can be immediately used by researchers to analyze the sources of learner difficulties. This sort of data makes it possible to investigate pervasive patterns of difficulty in the learners' English (that is, to investigate what some applied linguists call the 'interlanguage' of learners). Results of such analyses can directly enhance the entire web-based writing platform. Specifically, we have developed an authoring tool for designing online help which targets precisely the problems uncovered in the analysis of learners' errors. Moreover, based upon this sort of data, researchers can improve the design of teaching and reference materials. [See section 2.3 below for more details.]

2.2 The Student's View

A registered student logging onto the system is first shown a menu of links, including a link to a discussion board dedicated to the students in that class and links to helpful websites for ESL writers. To compose or turn in an essay, the student links to a page that displays a row of colored buttons, basically each button (or cluster of buttons) representing a different essay the student has written or is in the process of writing. From this page, the student can opt to resume work on an unfinished essay or revision, or to submit or compose a new essay [see Figure 2].

To compose an essay, students can elect either to compose online by typing their essay within a designated text box on the appropriate page or to copy and paste into that box an essay composed off-line. The latter essays are imported as text files.

From this screen where the essay has been composed or imported, students can submit the essay to the instructor. Alternatively, through a drop-down list of all of their classmates' names they can send the essay to any number of their classmates for

peer editing or commenting. The methods of selecting portions of text for comment and for submitting comments are basically the same as under the teacher's view described above.

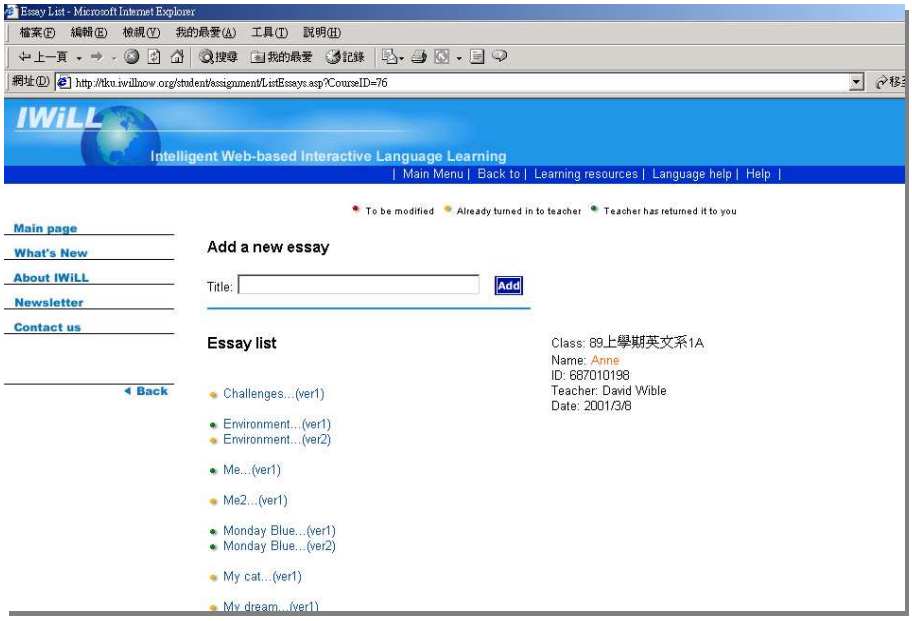

Figure 2: *Students' interface for turning in new essays or checking returned essays.*

When a student views an essay that has been marked by the teacher, the essay itself appears almost identical to the student's original, unmarked essay. None of the teacher's comments are immediately visible. The only difference in the appearance of the marked and unmarked version of an essay is that in the 'corrected' version some of the student's text shows up in blue. These are portions of the essay that the teacher has marked for comment. To see the content of the teacher's comment, the student places the cursor on the blue text and the comment appears.

An important feature offered to students is a specific sort of search function which they can access through a link labeled: "Search all comments in my essays." With this function, a student can access a list of all of the comments that the teacher has marked on his essays. The comments are listed in descending order of frequency as they occur in the entire set of that particular student's essays.

By clicking on the View button for any of these comments that appear on this display, the student retrieves a cumulative listing of all of the instances where this comment appeared in his own essays. To give the minimum context that would allow the student to see the nature of the marked problem, this search function retrieves complete sentences from the student's texts even if the teacher had marked only a word or phrase or other proper subpart of that sentence for comment. In instances where the teacher has marked off a chunk of text which spans a sentence boundary in the student essay, the entire text of both (or all) of those sentences is displayed. By

clicking on any of the tokens that have been retrieved, the student links to the complete text of the essay from which that token was extracted, thus accessing the full context.

What the "View Comments" function provides is the opportunity for the student to see patterns of difficulty, to see in one glance a set of tokens of one type of difficulty from his own writing. Of course, what is needed here is research on the differential effects of the two approaches to providing feedback. Moreover, the effectiveness of the View Comments function will almost certainly depend not simply on the fact that the system allows searches of the essays according to teacher comments, but also upon the quality and clarity of the comments themselves. An important property designed into the system is that it can track precisely the kinds of data needed for investigating these sorts of issues.

Another feature that exploits the system's record of teacher feedback to the learner is what we refer to as system-triggered help. This feature automatically tracks the number of times a teacher marks the same error type within the writing of a single student. For certain error types that have help pages available, such as sentence fragments, the system will automatically notify the student when the number of instances of that error type marked by the teacher in his or her essays exceeds a preset threshold. For example, if the teacher marks more than three sentence fragments in a student's writing, the system will notify that student with a message that includes a link to the help page on that type of error, in this case sentence fragments, and the help is offered in the students' first language, Chinese [see Figure 3]. The result is that the teacher need no longer shoulder the oppressively taxing task of keeping track of each learner's various areas of difficulty. The system tracks this and responds to the learners accordingly.

2.3 The Researcher's Tools

The system has been designed to create a corpus of student essays as a byproduct of the teacher-student interaction on the system. Specifically, each essay that a student submits to his teacher over this system is, with the permission of that student, copied into a corpus of "learner English." Consequently the corpus itself grows as the system is used by students and their teachers.

2.3.1 Automated Corpus Capture

The creation and analysis of corpora of learner language data is an extremely new and promising field of research [see Granger (98)]. One of the formidable obstacles in this field is a practical one of how to input the learner data. [Granger (98):11] mentions three methods, all extremely tedious, time-consuming, expensive a) nd the first two prone to error: (1) scanning essays from hard copies and (2) keying in data manually (3) downloading electronic data. Granger implies that the latter refers to collecting student essays that are on disks. Our system offers another way of creating learner corpora which goes a long way toward eliminating these prohibitive drawbacks. The texts created by students enter the corpus virtually unaffected by

any intermediate steps for "inputting" them because the exact text that the student sends to the teacher over the system is copied into the corpus. The task of creating the corpus is thus streamlined and reduces essentially to certain sorts of maintenance work to delete duplicate essays or revised versions of earlier essays.

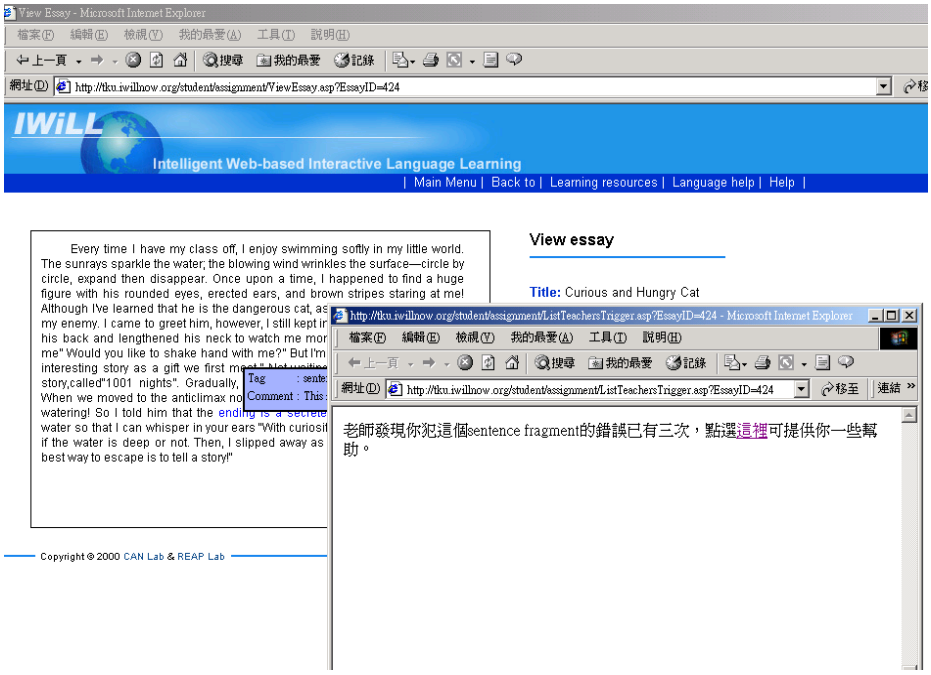

Figure 3: *Students' view of automatic teacher-triggered help given in learners' L1.*

Moreover, when students first register to use the system, they provide relevant metadata about their years of studying English, their gender, age, mother tongue, and the relevant fields of metadata are updated every semester. Each essay a student turns in is automatically indexed to this information and annotated with the date when that specific essay was submitted. This indexing allows for longitudinal studies of learner writing as well as cross-sectional studies that consider variables such as gender, age, or years of study. Researchers can add other fields of metadata to track other variables for specific studies. Researchers are not only able to search the corpus of essays collected from learners.

2.3.2 Interlanguage Analysis and the Creation of Online Grammar Help

The results of the researcher's analyses of learner difficulties can be translated into the content of an active online help function for those learners. The system includes an authoring environment for content administrators (ICPs) where they simply indicate what string of text in a learners essay should trigger help, and then write the content of the help which should be displayed for that particular string. Research on the learner corpus has revealed, for example, that the word 'ever' was misused by learners in 25% of the cases where it appeared in their essays. Further analysis attributed this to negative transfer in which learners associated the English expression 'ever' with a Chinese counterpart expression (*cheng jing*). These two expressions while overlapping in use and meaning, diverge in important ways, and it is precisely in these diverging respects where students misused the English expression. 'Ever' is what linguists refer to as a negative polarity item, noting that its distribution is restricted to contexts where it co-occurs with other expressions in particular configurations. Specifically, 'ever' is typically restricted to appearing in negative, interrogative, or conditional clauses. The Mandarin counterpart, however, has no such restriction. Hence, learners of English with a Chinese L1 commonly neglect the negative polarity constraints governing 'ever' and produce sentences like "I have ever read an article on that topic." Based on this linguistic analysis, the IWiLL authoring environment for online help was used to design advice concerning the word 'ever' addressing precisely the difficulties it poses for Chinese learners. When learners request general help on an essay, the help function actively detects instances of 'ever', highlights them and creates a link to this advise.

Another example of interlanguage analysis of the learner corpus which has led to automated online help is based on the learners' use of the expression "no matter." A keyword search of the learner corpus revealed that Taiwan's learners misused this expression in one third of the instances where it occurred. Moreover, a detailed look at the misused cases showed that learners had not yet acquired the constraint that "no matter" must be followed by an interrogative pronoun (who, what, when, where, etc.) or interrogative complementizer (whether or if). This stage of acquisition was attributed to the influence of L1 since the counterpart Chinese expression (*bu guan*) has no such restriction. Based upon these results, an automatic grammar checking component was designed to detect all misuses of "no matter" as in "No matter you are rich or poor, you need friends." The system responds by offering the learner an immediate link to a tutorial that explains the misleading differences between English "no matter" and Chinese *bu guan* and offers positive and negative examples for practice [see Figure 4].

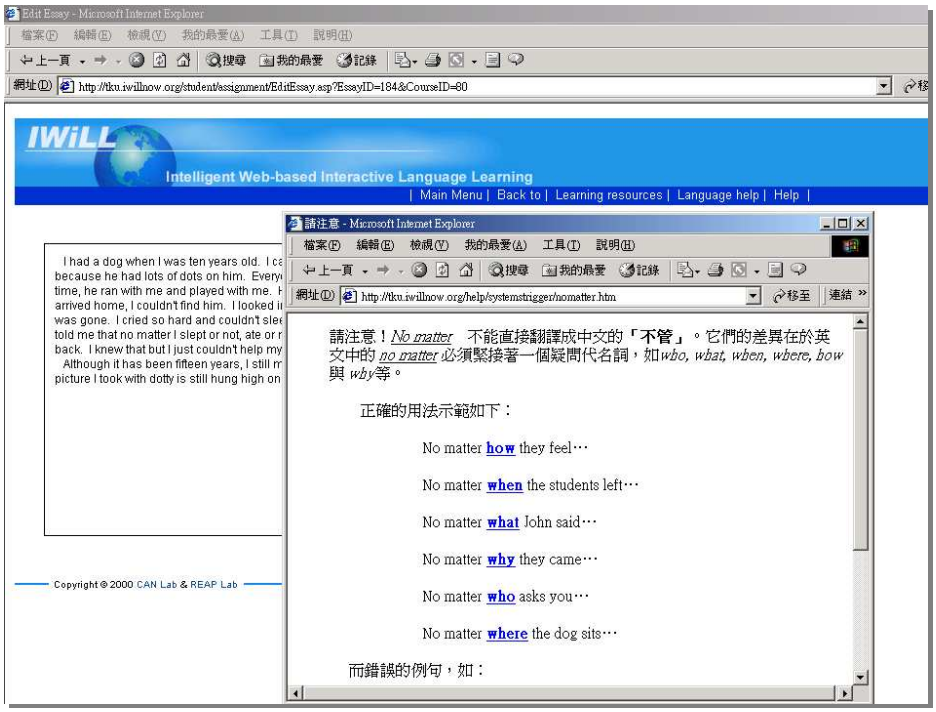

Figure 4: *Grammar help which automatically detects misuses of "no matter" and provides a tutorial pop-up*

2.3.3 Analysis of Interlanguage and L2 Pedagogical Practice

The platform not only archives a learner corpus; that corpus is annotated with teacher comments. Moreover, the system makes it possible to extract all strings in the student writing that have been marked with a particular error, for example, all instances within the corpus that have been marked as sentence fragments. This data has a wide variety of potential research applications. It allows, for example, comparison of different teachers' style of correction. It can create profiles of which error types teachers care most about. For example, the teacher-annotated corpus shows that by far the most common error type marked by teachers is word choice errors. By storing each instance of a teacher's comment on an essay and indexing each comment to the exact strings of student text marked with that comment, the system permits easy extraction of each token in the learner corpus that has been marked as a particular error type. With this feature, we were able not only to discover that word choice is by far the most common error type marked by teachers, we were able to extract every single token marked as a word choice error. Hand analysis of these tokens revealed that one in five of the word choice errors were collocation errors. These results have provided us with valuable data concerning which sorts of errors should be designed into a grammar checker. For example, the

discovery of the pervasiveness of collocation errors has led us to design an automatic collocation correction software within the IWiLL writing environment (called Lexical Assistant) which makes use of WordNet and standard English corpora to offer lexical selections as alternatives to mis-collocations [Wible et al (01)] [see Figure 5].

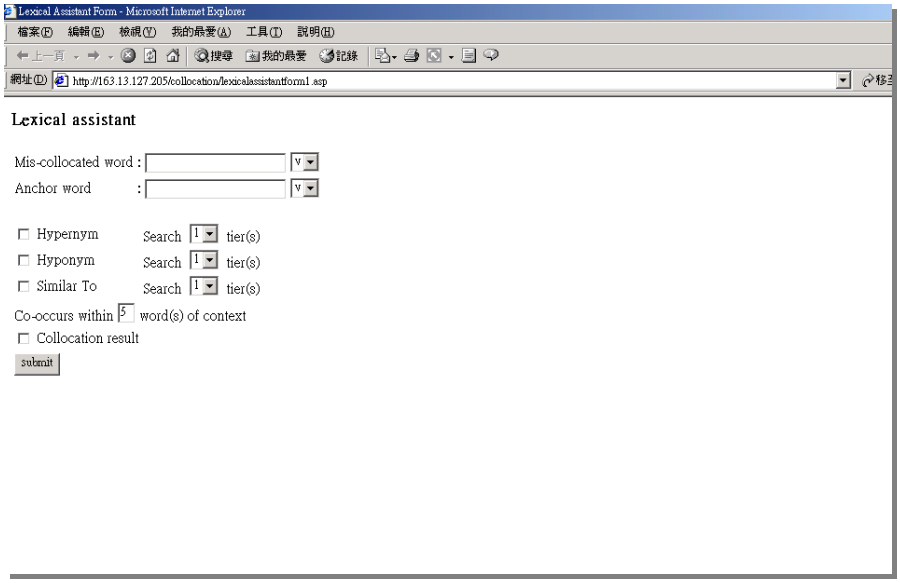

Figure 5: *Teachers' and researchers' interface for Lexical Assistant*

3 Conclusion

The underlying goal of the project described above has been not only to create an online writing environment that connects teachers and students by way of a user-friendly interface, but also to provide ways to exploit the valuable data that is created when the environment is used. The learners' essays themselves are stored in growing corpus of ESL language production. The comments that teachers append to the particular segments of the learners' texts in the course of essay correction are treated as annotations of those texts, which can be searched and retrieved. An authoring environment for online help permits content administrators to turn interlanguage research results into highly specific help concerning attested difficulties which traditional language education has neglected. It is hoped that increasingly sophisticated and dynamic manipulations of these sorts of data will lead to the delivery of evermore useful and useable information to learners, teachers, and researchers both online and off.

Acknowledgments
The research reported here was supported in part by a Pursuit of Excellence research grant from the Ministry of Education of the ROC, grant #89-H-FA07-1-4-2. This support is hereby gratefully acknowledged.

References

[Ferris (96)] Ferris, D.: "The case for grammar correction in L2 writing classes: a response to Truscott"; (1996). *Journal of Second Language Writing,* 8-1 (1999),1-11.

[Granger (98)] Granger, S.: "The computer learner corpus: a versatile new source of data in SLA research.", In Sylviane Granger, *Learner English on Computer*.; Longman, New York (1998).

[James (98)] James, C.: "Errors in Language Learning and Use."; Longman, New York (1998).

[Kearsley (89)] Kearsley, G.: "Computer-aided instruction, intelligent", In S. Shapiro (ed.) pp. 154-9, (1987).

[Kuo et al (00a)] Kuo, C.-H., David Wible, and Chia-Lin Chio,: "On the design of a synchronous virtual writing clinic."; (manuscript), (2000).

[Kuo et al (00b)] Kuo, C.-H., David Wible, Meng-Chang Chen, Li-Chun Sung, Nai-Lung Tsao, Chia-Lin Chio, and Chih-Chiang Wang: "The design and implementation of an intelligent web-based interactive language learning system." (manuscript), (2000).

[Krashen (94)] Krashen, S.: "The input hypothesis and its rivals", in Ellis, N. (ed.) *Implicit and Explicit Learning of Languages*; Academic Press, London (1994).

[Nerbonne (98)] Nerbonne, J. (ed): "*Linguistic Databases*. Stanford"; CSLI (Center for the Study of Language and Information) (1998).

[Neuwirth et al. (96)] Neuwirth, M.C., Wojahn, G.P.: "Learning to write: computer support for a cooperative process.", In T. Koshmann, CSCL: Theory and Practice of an Emerging Paradigm. Lawrence Erlbaum Associates; (1996), pp.147-70.

[Ogata et al. (98)] Ogata, Hiroaki, Yano, Yoneo, and Wakita, Riko CCML — "Exchanging marked-up documents in a networked writing classroom."; CALL (*Computer Assisted Language Learning),* 11-2 (1998), 201-214.

[Selinker (92)] Selinker, L.: "*Rediscovering interlanguage*."; Longman, London (1992).

[Selinker (96)] Selinker, L.: "On the notion 'IL competence' in early SLA research: an aid to understanding some baffling current issues.", In Brown, G., Malmkjaer, K. & Williams, J. (eds.). *Performance and competence in second language acquisition*.; Cambridge University Press, Cambridge, (1996),92-113.

[Truscott (96)] Truscott, J.: "The case against grammar correction in L2 writing classes."; *Language Learning,* 46-2 (1996),327-369.

[Wible et al. (01)] Wible, D., Chin-Hwa Kuo, Nai-lung Tsao and Anne Liu: "An Investigation in Using WordNet to Automate the Correction of Language Learners' Collocation Errors" Unpublished manuscript, Tamkang

University Taipei, Taiwan.

[Wible et al. (00)] Wible, D., Chin-Hwa Kuo, Nai-lung Tsao, Anne Liu, Li-Chun Sung, and Chia-Lin Chio,: "Putting learners first: an integrated multimedia environment for language learning" International Conference on Engineering Education, Taipei (2000).

[Wible] Wible, D.: "SLA research, second language writing, and the role of web-based learning environments," Unpublished manuscript, Tamkang University, Taipei, Taiwan.