

# **Evaluating Two Web-based Grammar Checkers - Microsoft ESL Assistant and NTNU Statistical Grammar Checker**

**Hao-Jan Howard Chen\***

## **Abstract**

Many ESL students need to improve writing skills to pass various language tests; thus, writing teachers need to read many compositions and provide feedback. To help ESL teachers reduce their teaching load and to give students faster feedback, various English grammar checkers have been developed. Few of these PC-based grammar checkers, however, are widely available to ESL learners. As the Internet has become an important tool for language education, web-based grammar checkers have begun to emerge. In this paper, we first introduce two new web-based grammar checkers (Microsoft ESL Assistant and NTNU statistical grammar checker) and then compare their performance. Ten common EFL errors selected from a large Chinese EFL learner corpus were used to test these two grammar checkers. The test results showed that the NTNU statistical checker was far more sensitive to various learner errors, and it could detect eight types of selected errors. Microsoft ESL Assistant could only deal with five types of errors. Moreover, these two checkers both could not deal with fragments and run-on sentences errors. It seems clear that both checkers have room for improvement before they can provide satisfactory service to ESL learners. The Microsoft ESL Assistant should expand its coverage to detect more learner errors. NTNU checker should reduce false alarms and indicate the locations of errors more accurately. Learner errors are indeed complicated for developers of grammar checkers, but the strong need for a functional grammar checker deserves CALL researchers' special attention.

**Keywords:** ESL Writing, Errors, Grammar Checker, Ngrams, Rules.

---

\* English Department, National Taiwan Normal University  
E-mail: hjchen@ntnu.edu.tw

## 1. Introduction

It is challenging for second language learners to become proficient writers of the target language. Learners need considerable writing practices before they can write accurately and fluently. In addition to providing more writing opportunities to students, many teachers and researchers believe that learners need to receive proper corrective feedback on their writings (Ferris, 1999, 2003, 2006). If learners only keep on writing and do not receive any corrective feedback, they will not be able to make progress quickly. Even though the role of corrective feedback in second language learning remains controversial, many teachers and learners firmly believe that feedback plays an important role in second language writing (*cf.* Ferris, 1999, 2003, 2006; Truscott, 1996, 1999, 2007; Goldstein, 2006; Guenette, 2007).

In many ESL/EFL settings, writing teachers need to work with 40 or 50 students in their classes. Reading and correcting students' essays is a great burden for many writing teachers. To reduce teachers' loads in correcting common errors and to help students enhance their writing accuracy, various grammar checking tools have been developed in different countries. Many CALL researchers consider the development of grammar checkers to be part of the Intelligent CALL research (*cf.* Holland *et al.*, 1995). The development of a useful grammar checker to identify and correct learners' errors has been considered a very important research direction in CALL. However, most of the English grammar checkers developed by academic institutes could only deal with a limited set of grammar errors and were not made available to ESL students (*e.g.*, Liou, 1991, 1992, 1993; More, 2006; Naber, 2003; Park, Palmer, & Washburn, 1997). Some commercial PC-based grammar checkers (*e.g.* Whitesmoke) are available, but their error detecting capacities are still limited according to some recent evaluation studies (Chiu, 2008).

With the rapid development of artificial intelligence and natural language processing technologies, several automated essay scoring programs (*e.g.*, Vantage My Access and ETS Criterion) have appeared recently in the ESL market (Attali, 2004; Burstein, Chodorow & Leacock, 2004; Elliot, 2001; Elliot & Mikulas, 2004; Han, Chodorow & Leacock, 2006; Higgins, Burstein & Attali, 2006). Grammar checkers are also included in these two leading writing tools. By subscribing to these commercial programs, students can choose from a wide range of practice essays topics to write multiple drafts and receive immediate corrective feedback in the form of both holistic scores and diagnostic comments on grammar, organization, style, and usage.

These commercial software packages, however, are expensive, and ESL students who subscribe to these services can only use these programs for 3-6 months during the subscription period. Because of these limitations and high prices, many ESL/EFL students still do not have access to any of these programs. The better way of helping a large number of students is to

look for similar grammar checking programs on the Internet. Although the Internet has become one of the most important resources for second/foreign language education worldwide, very few web sites offer grammar checking tools for ESL learners.

There are some reasons for the limited availability of grammar checkers. The most important reason is that it is very difficult to develop a reliable grammar checker. Daniel Kies, on his web site (<http://papyr.com/hypertextbooks/grammar/gramchek.htm>), compared several popular grammar checking tools. These PC tools are Microsoft's Word (both the Windows and the Mac versions of the program), Corel's WordPerfect (Windows only), Grammarian Pro X (Mac only), and Open Office Writer with the Language Tool extension added (platform independent). The best program for Windows is Corel's WordPerfect, but it can only reach about 40 percent accuracy for about twenty types of errors.

According to Naber (2003), there are basically three different ways to implement a grammar checker. Each approach has its strengths and weakness.

1. Syntax-based checkers. In this approach, a text is completely parsed, *i.e.* the sentences are first analyzed and each sentence is assigned a tree structure. The text is considered incorrect if the parsing does not succeed. A robust syntactic parser plays a very important role in this approach (*cf.* Jensen, 1993).
2. Statistics-based checkers. In this approach, a language model is trained from a large training corpus (*e.g.*, a native corpus), which contains many short phrases (*cf.* Atwell & Elliott, 1987; Chodorow & Leacock, 2000). It can be used for detecting and correcting certain types of grammar errors, where local information is sufficient to make decision. Sequences which occur often in the corpus can be considered correct in other texts, and uncommon sequences might be errors. Some recent developments of grammar checkers based on corpus linguistics and statistics seem to be quite useful (Chodorow & Leacock, 2000; Sjobergh, 2005; Wu, Su, Jiang, & Hsu, 2006). A popular example is the grammar and style checker developed by ETS.
3. Rule-based checkers. In this approach, a set of rules is matched against a text which has at least been POS tagged. This approach is similar to the statistical approach, but all the grammar rules are developed manually. Park *et al.* (1997) and Naber (2003) are studies using this approach.

Although it is time-consuming and difficult to develop a robust grammar checker that can provide specific and clear grammar feedback, many ESL/EFL teachers and students worldwide desperately need a good grammar checker to improve teaching and learning. With funding from the National Science Council of Taiwan, a research team at National Taiwan Normal University (NTNU) developed a statistical grammar checker based on large English corpora. At the same time, another research team in Microsoft also developed their new grammar checker called Microsoft ESL Assistant. It seems that a grammar checking service has begun to emerge on the Web. These two projects are briefly introduced in the sections below.

## 1.1 Microsoft ESL Assistant

The Microsoft Research ESL Assistant is a web service that provides correction suggestions for ESL writing errors. The web service also provides suggestions for word choice. It consists of three parts: a set of modules that identify possible corrections, a large language model that evaluates the possible suggestions, and a module that produces search results using Live Search. The individual error modules each target specific errors. Some of these models are based on heuristics, while others use machine learned classifiers. Information that the modules take into account includes the presence of specific words as well as the sequence of part-of-speech tags that are automatically assigned. The language model of ESL Assistant is trained on the Gigaword corpus. The English Gigaword Corpus is a comprehensive archive of newswire text data that has been acquired over several years by the Linguistic Data Consortium (LDC) at the University of Pennsylvania.

The ESL Assistant has recently been updated with a new user interface that uses the Microsoft Silverlight™ browser plug-in. As shown below in Figure 1, text to be checked is entered in the box at the top. When the user clicks the check button, potential errors are identified and highlighted. Users can proceed from one highlighted segment to the next, reviewing the possible errors and the suggestions presented in the box below. In addition, a pie chart allows the user to compare the approximate frequency distributions of their own input and the suggestions, as found by the Microsoft Bing™ search engine. When the user hover their mouse over any of the options available, a dropdown panel shows selected usage examples found on the Web. Users can explore additional examples by clicking the link at the bottom of the dropdown panel.

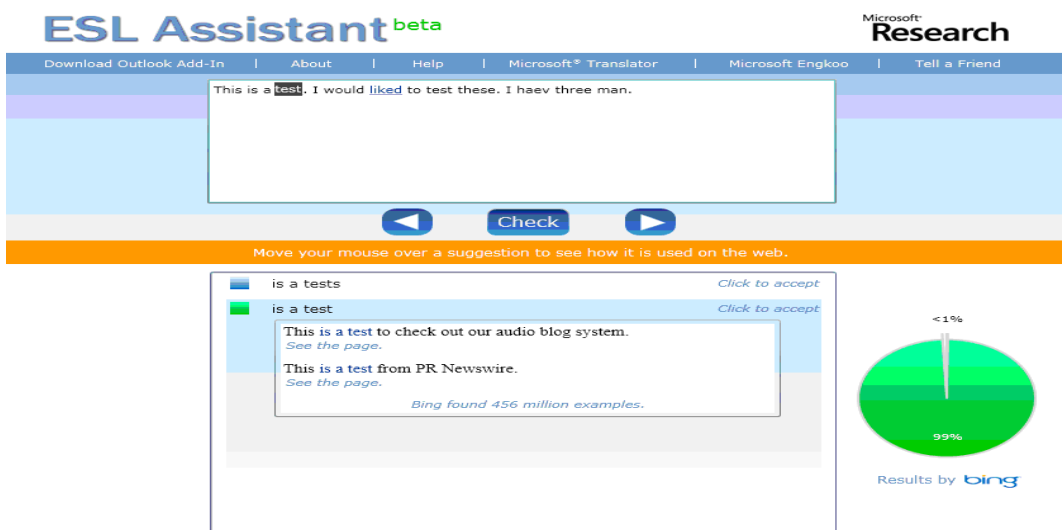


Figure 1. The Microsoft ESL assistant.

## 1.2 NTNU Statistical Grammar Checker

The ngram-based statistical grammar checker was developed at NTNU. The grammar checker was based on the statistical approach. This is the approach currently used by ETS and several other research teams (e.g., Chodorow & Leacock, 2000; Sjobergh, 2005). The advantages of using an ngram-based statistical approach follow. First, there is no need to write grammar rules manually. Second, it can detect many errors which cannot be detected by traditional grammar checkers and parsers. The bigrams and trigrams can detect more errors. Third, the lexical errors are more likely to be detected by this type of grammar checker.

The procedures used to detect the violations of English grammar in a statistical grammar checker are not complicated. The grammar checker system is first trained on a large corpus of edited texts, from which it extracts and counts *bigrams* that consist of sequences of adjacent words. British National Corpus (BNC) was used as the reference corpus for the NTNU checker. The British National Corpus is a 100-million-word collection of samples of written and spoken English from a wide range of sources. The corpus covers British English of the late twentieth century from a wide variety of genres with the intention that it be a representative sample of spoken and written British English of that time. Then SRI (Stanford Research Institute) Language Modeling Toolkit (SRILM) was used to develop an ngram language model based on the BNC corpus. Various bigrams and trigrams were extracted from the BNC corpus and stored in a large database. A web-based grammar checker interface linked to the ngram database was also developed. The web interface of the ngram-based grammar checker is shown in Figure 2.



Figure 2. The Web interface of NTNU Ngram-based statistical grammar checker.

ESL students can submit their writings to the web-based grammar checker, and the system searches the students' writings for bigrams and trigrams. The bigrams and trigrams which never or rarely show up in BNC corpus were highlighted by the system. Within a few seconds, this checker can quickly check the article and highlight the problematic word strings (clusters) not found in the native corpora. With the help of the bigram and trigram information from the very large native corpora, the system can efficiently detect and highlight the problematic usage in students' essays.

After students input their essays into the grammar checker, the potential errors are automatically highlighted in red. The Google search engine can be used to search for the better usage when users use the mouse to highlight any word string in their writing. Google either directly recommends some more commonly used strings or shows various sentences which include these words.

Since the Microsoft ESL Assistant and NTNU ngram-based statistical checker are web-based services, they have the potential to allow many ESL students to use them. It is not clear, however, if these two new checkers can reach the accuracy level of some of the leading products in detecting or correcting ESL student errors. To determine their usability, it is essential to further evaluate their performance in dealing with ESL learner errors.

## 2. Methodology

To assess if these two grammar checkers can provide similar services as the leading products in the market, it is necessary to compare the performance of these two checkers with some of the leading commercial products in the market. One of the best grammar checking tools available is the ETS Criterion. It can detect many ESL learner errors and can provide appropriate feedback. Based on the results of a previous research evaluation study (Chen, 2009), ten types of common errors, as shown below in Table 1, which can be correctly identified by ETS Criterion were used to test the performance of the Microsoft ESL Assistant and NTNU grammar checker. Five sentences in each error category were randomly selected from a large set of incorrect sentences which were identified by ETS criterion.

***Table 1. Ten types of common ESL errors detected by ETS Criterion.***

Error Types of ETS Criterion
1. Missing or Extra Article
2. Spelling
3. Fragment or Missing Comma
4. Run-on Sentences
5. Subject-Verb Agreement
6. Confused Words
7. Ill-formed Verbs
8. Proofread This!
9. Wrong Article
10. Compound Words

Three research questions were proposed in this paper.

1. What types of errors detected by ETS Criterion can be detected by these two web-based grammar checkers?
2. What types of errors cannot be identified by these two grammar checkers?
3. What are the accuracy rates of these two grammar checkers? Which checker has the higher accuracy rate? Which checker can find more errors?

### 3. Evaluation Result

#### 3.1 The Performances of Two Grammar Checkers in Each Type of Error

In the following section, the performances of each grammar checker in each different error category are shown in tables. The errors detected by each grammar checker are underlined. The first part of the table shows the errors marked by ETS Criterion, the second part shows the errors marked by Microsoft ESL Assistant, and the third part shows the errors marked by the NTNU ngram-based grammar checker.

1. Missing or Extra Article: As shown in Table 2, Microsoft ESL Assistant could detect 4 out of 5 errors for missing or extra articles. The NTNU grammar checker could also detect 4 errors, but the highlighting of errors was not accurate and there were some false alarms. Although both checkers could detect this type of errors, it was clear that ESL Assistant was more effective.

**Table 2. Comparison of checker performance on missing or extra articles.**

<b>ETS Criterion</b>	<ol style="list-style-type: none"> <li>1. From going to <u>lab</u> to do experiments, I learned a lot.</li> <li>2. Here is an old lady who speaks some words she sees on <u>street</u> randomly.</li> <li>3. In <u>first</u> grad of high school, I did not know anyone.</li> <li>4. It is <u>honor</u> that I am one of her best friends.</li> <li>5. Bear could get <u>good</u> relationship with everyone in our class.</li> </ol>
<b>Microsoft ESL Assistant</b>	<ol style="list-style-type: none"> <li>1. From going to lab to do experiments, I learned a lot.</li> <li>2. Here is an old lady who speaks some words she sees on <u>street</u> randomly.</li> <li>3. In <u>first</u> grad of high school, I did not know anyone.</li> <li>4. It is <u>honor</u> that I am one of her best friends.</li> <li>5. Bear could get <u>good</u> relationship with everyone in our class.</li> </ol>
<b>NTNU Grammar Checker</b>	<ol style="list-style-type: none"> <li>1. From going to <u>lab to do experiments</u>, I learned a lot.</li> <li>2. Here is an old lady who <u>speaks some words she sees on street</u> randomly.</li> <li>3. In first <u>grad of</u> high school, I did not know <u>anyone</u>.</li> <li>4. It is <u>honor</u> that I am one of her best friends.</li> <li>5. Bear could <u>get</u> good relationship with everyone in <u>our class</u>.</li> </ol>

2. Spelling errors: As shown in Table 3, Microsoft ESL Assistant could not detect any spelling errors. The NTNU checker, however, could detect all the 5 spelling errors. It was clear that the NTNU checker was more effective in this category. It is odd that why Microsoft ESL Assistant did not incorporate the very effective Microsoft spelling checker into the web system.

**Table 3. Comparison of checker performance on spelling errors.**

<b>ETS Criterion</b>	<ol style="list-style-type: none"> <li>1. I am a <u>deligent</u> student.</li> <li>2. She hated to study the textbook and even <u>frquently</u> skipped the class.</li> <li>3. but it <u>seemes</u> everyone is so busy.</li> <li>4. She is also <u>humurous</u>.</li> <li>5. He is very hansome.</li> </ol>
<b>Microsoft ESL Assistant</b>	<ol style="list-style-type: none"> <li>1. I am a deligent student.</li> <li>2. She hated to study the <u>textbook</u> and even frquently skipped the class.</li> <li>3. But it seemes everyone is so busy.</li> <li>4. She is also humurous.</li> <li>5. He is very hansome.</li> </ol>
<b>NTNU Grammar Checker</b>	<ol style="list-style-type: none"> <li>1. I am a <u>deligent</u> student.</li> <li>2. She hated to study the <u>textbook and even frquently skipped</u> the class.</li> <li>3. But <u>it seemes everyone</u> is so busy.</li> <li>4. She <u>is also humurous</u>.</li> <li>5. He is <u>very hansome</u>.</li> </ol>

3. Fragment or Missing Comma: As shown in Table 4, Microsoft ESL Assistant again could not detect any spelling errors. Similarly, the NTNU checker could not detect any of the fragment errors but it highlighted some words. It was clear that both types of grammar checkers could not deal with fragment errors.

**Table 4. Comparison of checker performance on fragment or missing comma.**

<b>ETS Criterion</b>	<ol style="list-style-type: none"> <li>1. <u>Although now we are in various areas over this island.</u></li> <li>2. <u>Because his or her friends are someone who relate to him or her and know about him.</u></li> <li>3. <u>When our friends' or family's birthday is coming.</u></li> <li>4. <u>In a word, because people analyze the question in different ways, such as the TV station and audience.</u></li> <li>5. <u>If she can be more considerate.</u></li> </ol>
<b>Microsoft ESL Assistant</b>	<ol style="list-style-type: none"> <li>1. Although now we are in various areas over this island.</li> <li>2. Because his or her friends are someone who relate to him or her and know about him.</li> <li>3. When our friends' or family's birthday is coming.</li> <li>4. In a word, because people analyze the question in different ways, such as the TV station and audience.</li> <li>5. If she can be more considerate.</li> </ol>



*Microsoft ESL Assistant and NTNU Statistical Grammar Checker*

<b>NTNU Grammar Checker</b>	<ol style="list-style-type: none"> <li>1. Although now we are in various areas <u>over</u> this island.</li> <li>2. Because his or her <u>friends are someone who relate</u> to him or her and know about him.</li> <li>3. When our <u>friends' or family's birthday</u> is coming.</li> <li>4. In a word, because people <u>analyze</u> the question in different ways, such as the TV station and audience.</li> <li>5. If she can be more <u>considerate</u>.</li> </ol>
-------------------------------------	---

4. Run-on Sentences: As shown in Table 5, Microsoft ESL Assistant could not detect any run-on sentence errors. Similarly, the NTNU checker again could not detect any of these run-on sentence errors and it also generated some false alarms. It was clear that both types of grammar checkers could not deal with run-on sentence errors.

**Table 5. Comparison of checker performance on run-on sentences.**

<b>ETS Criterion</b>	<ol style="list-style-type: none"> <li>1. <u>Like I would play strong and tell her firmly that she should go to Purdue no matter what and that's a chance of a life time, though in reality, and actually the first instinct, I forbid her to go.</u></li> <li>2. <u>Then the day came, Bob received the grading and bounced to Charlie.</u></li> <li>3. <u>Bob was turned down and felt disappointed and, however, felt angry at Charlie.</u></li> <li>4. <u>We have people and place, the rest is food.</u></li> <li>5. <u>It was like a dream, I was flying!</u></li> </ol>
<b>Microsoft ESL Assistant</b>	<ol style="list-style-type: none"> <li>1. Like I would play strong and tell her firmly that she should go to Purdue no matter what and that's a chance of a life time, though in reality, and actually the first instinct, I forbid her to go.</li> <li>2. Then the day came, Bob received the grading and bounced to Charlie.</li> <li>3. Bob was turned down and felt disappointed and, however, felt angry at Charlie.</li> <li>4. We have people and place, the rest is food.</li> <li>5. It was like a dream, I was flying!</li> </ol>
<b>NTNU Grammar Checker</b>	<ol style="list-style-type: none"> <li>1. Like I would play <u>strong</u> and tell her <u>firmly</u> that she should go to <u>Purdue</u> no matter what and <u>that's</u> a chance of a <u>life time</u>, though in reality, and actually the first instinct, I <u>forbid</u> her to go.</li> <li>2. Then the day came, <u>Bob received the grading and bounced</u> to Charlie.</li> <li>3. Bob was turned down and <u>felt disappointed</u> and, however, felt <u>angry</u> at Charlie.</li> <li>4. We have <u>people and place</u>, the rest <u>is food</u>.</li> <li>5. It was like a dream, I <u>was flying!</u></li> </ol>

5. Subject verb agreement: As shown in Table 6, Microsoft ESL Assistant could only detect one agreement error among the five errors. The NTNU checker, however, could detect five agreement errors. It was clear that the NTNU grammar checker performed much better in this category.

**Table 6. Comparison of checker performance on subject-verb agreement.**

<b>ETS Criterion</b>	<ol style="list-style-type: none"> <li>1. <u>There are</u> many advertisement.</li> <li>2. My <u>teacher tell</u> me a story.</li> <li>3. No matter what <u>I plans</u> after graduation.</li> <li>4. The <u>advertisement appear</u> in TV.</li> <li>5. <u>She sing</u> very loudly.</li> </ol>
<b>Microsoft ESL Assistant</b>	<ol style="list-style-type: none"> <li>1. There are many <u>advertisement.</u></li> <li>2. My teacher tell me a story.</li> <li>3. No matter what I plans after graduation</li> <li>4. The advertisement appear in TV</li> <li>5. She sing very loudly.</li> </ol>
<b>NTNU Grammar Checker</b>	<ol style="list-style-type: none"> <li>1. There are <u>many advertisement.</u></li> <li>2. My <u>teacher tell</u> me a story.</li> <li>3. No matter what <u>I plans</u> after graduation.</li> <li>4. The <u>advertisement appear</u> in TV.</li> <li>5. She <u>sing very</u> loudly.</li> </ol>

6. Confused Words: As shown in Table 7, Microsoft ESL Assistant could only detect three errors which were related to the confusion between *a* and *an*. Also, it could not detect any confused word pairs (quite vs. quiet; effect vs. affect). The NTNU checker could detect the five confused words errors, although the error highlighting is not very accurate. It was clear that NTNU ngram grammar checker could find more errors in this category.

**Table 7. Comparison of checker performance on confused words.**

<b>ETS Criterion</b>	<ol style="list-style-type: none"> <li>1. It should be <u>an</u> perfect idea.</li> <li>2. I was <u>a</u> ambassador sent by the Chin dynasty.</li> <li>3. She is more independent as <u>a</u> individual.</li> <li>4. He is a very <u>quite</u> student.</li> <li>5. The new drug might have a better <u>affect</u>.</li> </ol>
<b>Microsoft ESL Assistant</b>	<ol style="list-style-type: none"> <li>1. It should be <u>an</u> perfect idea.</li> <li>2. I was <u>a</u> ambassador sent by the Chin dynasty.</li> <li>3. She is more independent as <u>a</u> individual.</li> <li>4. He is a very quite student.</li> <li>5. The new drug might have a better affect.</li> </ol>
<b>NTNU Grammar Checker</b>	<ol style="list-style-type: none"> <li>1. It should be an <u>perfect</u> idea.</li> <li>2. I was a <u>ambassador</u> sent by the Chin dynasty.</li> <li>3. She is more <u>independent as a</u> individual.</li> <li>4. He is a very <u>quite</u> student.</li> <li>5. The new drug might have a <u>better affect</u>.</li> </ol>

7. Ill-formed Verbs: As shown in Table 8, Microsoft ESL Assistant could only detect one of the five verb form errors. The NTNU checker could detect three errors correctly but the error highlighting does not indicate the error location very accurately. It seems that NTNU ngram grammar checker performed better in this category.

**Table 8. Comparison of checker performance on ill-formed verbs.**

<b>ETS Criterion</b>	1. I like the lessons they <u>have gave</u> us. 2. I should <u>have react</u> like Washington when I face the situation. 3. We <u>are excel</u> at different subject. 4. Please be sure that they <u>are satisfy</u> with it. 5. She and me are so happy <u>to had</u> a friend like this.
<b>Microsoft ESL Assistant</b>	1. I like the lessons they <u>have gave</u> us. 2. I should have react like Washington when I face the situation. 3. We are excel <u>at</u> different subject. 4. Please be sure that they are satisfy with it. 5. She and me are so happy to had a friend like this.
<b>NTNU grammar Checker</b>	1. I like the <u>lessons they have</u> gave us. 2. I should have <u>react like Washington when I face</u> the situation. 3. We are <u>excel at different</u> subject. 4. <u>Please be sure</u> that they are <u>satisfy</u> with it. 5. She and <u>me are</u> so happy to had a friend like this.

8. Proofread This!: As shown in Table 9, Microsoft ESL Assistant could detect three out of the five “proofread this” errors. The NTNU checker could detect all the five errors correctly though it also provided one false alarm, “player”. It was clear that the NTNU ngram grammar checker still performed better in this category.

**Table 9. Comparison of checker performances on proofread this!**

<b>ETS Criterion</b>	1. She had a <u>warn</u> family. 2. <u>I mayor</u> in engineering at school. 3. I'm a fourth <u>years student</u> of National Taiwan Normal University. 4. I find a job that <u>I desirable</u> . 5. The famous player kicked the ball at the time <u>they bended!</u>
<b>Microsoft ESL Assistant</b>	1. She had a <u>warn</u> family. 2. I mayor in engineering at school. 3. I'm a fourth <u>years</u> student of <u>National</u> Taiwan Normal University. 4. I find a job that I desirable. 5. The famous player kicked the ball at the time they <u>bended!</u>
<b>NTNU grammar Checker</b>	1. She had a <u>warn</u> family. 2. I <u>mayor in engineering</u> at school. 3. I'm <u>a fourth years student of National Taiwan Normal University</u> . 4. I find a job that <u>I desirable</u> . 5. The famous <u>player</u> kicked the ball at the time <u>they bended!</u>

9. Wrong Article: As shown in Table 10, Microsoft ESL Assistant could detect three out of the five article errors and generated one false alarm. The NTNU checker also could identify four errors correctly, but it also provided more false alarms. It was clear that the NTNU ngram grammar checker still performed better in this category.

**Table 10. Comparison of checker performance on wrong article**

<b>ETS Criterion</b>	<ol style="list-style-type: none"> <li>1. I turned <u>these</u> experience into my own energy for next challenge.</li> <li>2. <u>Those</u> stuff in high school days is absolutely insufficient.</li> <li>3. People always like to be <u>a</u> popular guys.</li> <li>4. Finding <u>a</u> best friend can make your life more special.</li> <li>5. If you do not have <u>a</u> slightest idea.</li> </ol>
<b>Microsoft ESL Assistant</b>	<ol style="list-style-type: none"> <li>1. I turned these <u>experience</u> into my own energy for <u>next</u> challenge.</li> <li>2. Those stuff in high school days is absolutely insufficient.</li> <li>3. People always like to be a popular <u>guys</u>.</li> <li>4. Finding a best friend can make your life more special.</li> <li>5. If you do not have <u>a slightest</u> idea.</li> </ol>
<b>NTNU Grammar Checker</b>	<ol style="list-style-type: none"> <li>1. I turned <u>these experience</u> into my own <u>energy for next</u> challenge.</li> <li>2. Those stuff <u>in high school days is absolutely</u> insufficient.</li> <li>3. <u>People always like</u> to be a <u>popular guys</u>.</li> <li>4. Finding a <u>best friend</u> can make your <u>life more</u> special.</li> <li>5. If you do not have a <u>slightest</u> idea.</li> </ol>

10. Compound Words: As shown in Table 11, Microsoft ESL Assistant could not detect any of the five compound word errors. The NTNU grammar checker could identify four compound word errors correctly, but it also generated a few false alarms. Although the error highlighting is not very accurate, users can highlight any word strings and get useful feedback from the Google search engine. It was clear that the NTNU ngram grammar checker still performed better in this category.

**Table 11. Comparison of checker performance on compound words**

<b>ETS Criterion</b>	<ol style="list-style-type: none"> <li>1. <u>When ever</u> I feel lonely.</li> <li>2. Do no say <u>any thing</u> when you don't now what you are saying.</li> <li>3. <u>Every one</u> should have the experience of making friends with others.</li> <li>4. I try to find <u>any one</u> to substitute but it seems everyone is so busy.</li> <li>5. At last, <u>no body</u> will get benefits from cheating.</li> </ol>
<b>Microsoft ESL Assistant</b>	<ol style="list-style-type: none"> <li>1. When ever I feel lonely.</li> <li>2. Do no say any thing when you don't now what you are saying.</li> <li>3. Every one should have the experience of making friends with others.</li> <li>4. I try to find any one to substitute but it seems everyone is so busy.</li> <li>5. At last, no body will get benefits from cheating.</li> </ol>

<b>NTNU Grammar Checker</b>	1. When ever <u>I feel</u> lonely.
	2. Do no <u>say any thing</u> when you don't now what you are saying.
	3. Every <u>one should</u> have the experience of making <u>friends</u> with others.
	4. I try to find any <u>one to substitute</u> but it seems <u>everyone</u> is so busy.
	5. At last, <u>no body will get benefits from</u> cheating.

### 3.2 General Comments on the Performance of Grammar Checkers

Based on the detailed analysis of the ten common ESL errors, it was obvious that the NTNU ngram grammar checker performed better than the Microsoft ESL Assistant. The NTNU checker performed better in 7 error categories. The Microsoft checker only performed better in the “missing or extra article” category. In addition, both grammar checkers failed to detect any fragment and run-on sentence errors. Table 12 summarizes the results of comparing these two grammar checkers on the error types they can detect.

**Table 12. Summary of the performance of two grammar checkers**

Error types	Microsoft	NTNU
1. Missing or Extra Article	O (Better)	O
2. Spelling	X	O (Better)
3. Fragment or Missing Comma	X	X
4. Run-on Sentences	X	X
5. Subject-Verb Agreement	O	O (Better)
6. Confused Words	O	O (Better)
7. Ill-formed Verbs	O	O (Better)
8. Proofread This!	O	O (Better)
9. Wrong Article	O	O
10. Compound Words	X	O (Better)

Notes: O means that the checker can detect the errors; X means that the checker fails to detect the errors.

Another possible way to compare the performances of these two grammar checkers is to compare the precision and the recall rates of these two checkers. Precision and recall are two widely used measures for evaluating the performance of information retrieval systems. In information retrieval, a perfect Precision score of 1.0 means that every result retrieved by a search was relevant (but says nothing about whether all relevant documents were retrieved) whereas a perfect Recall score of 1.0 means that all relevant documents were retrieved by the search (but says nothing about how many irrelevant documents were also retrieved). In evaluating the performances of grammar checkers, a perfect precision score means that all the

errors found by a checker are indeed real errors. A perfect recall score means that a checker would find all the errors made by language learners.

Table 13 summarizes the precision and the recall rates of these two grammar checkers. It seems that Microsoft ESL Assistant and NTNU grammar checker have similar precision rates (50% vs. 61%). However, the recall rate of NTNU grammar checker is clearly higher than that of Microsoft ESL Assistant (72% vs. 30%). The results indicate that NTNU can detect more errors in ESL learners' writing.

**Table 13. Summary of the precision and recall rate of two grammar checkers**

	Microsoft		NTNU	
	<u>Precision</u>	<u>Recall</u>	<u>Precision</u>	<u>Recall</u>
1. Missing or Extra Article	4/4 (100%)	4/5 (80%)	4/7 (57%)	4/5 (80%)
2. Spelling	0 (0%)	0/5 (0%)	5/5 (100%)	5/5 (100%)
3. Fragment or Missing Comma	0 (0%)	0/5 (0%)	0 (0%)	0/5 (0%)
4. Run-on Sentences	0 (0%)	0/5 (0%)	0 (0%)	0/5 (0%)
5. Subject-Verb Agreement	1/1 (100%)	1/5 (20%)	5/5 (100%)	5/5 (100%)
6. Confused Words	3/3 (100%)	3/5 (60%)	5/5 (100%)	5/5 (100%)
7. Ill-formed Verbs	1/2 (50%)	1/5 (20%)	4/6 (67%)	4/5 (80%)
8. Proofread This!	3/4 (75%)	3/5 (60%)	5/6 (83%)	5/5 (100%)
9. Wrong Article	3/4 (75%)	3/5 (60%)	4/8 (50%)	4/5 (80%)
10. Compound Words	0 (0%)	0/5 (0%)	4/7 (57%)	4/5 (80%)
<b>Average</b>	50%	30%	61 %	72%

#### 4. Discussion

Based on the test results, eight types of errors identified by ETS Criterion can also be detected by the NTNU grammar checker. Microsoft ESL Assistant can only detect five types of learner errors. It seemed obvious that the NTNU grammar checker is more sensitive to ESL learner errors. The Microsoft ESL Assistant seems less sensitive in reporting learners' errors. Some common ESL errors, such as spelling and subject-verb agreement, are not treated by the current version of Microsoft ESL Assistant. It is unclear why Microsoft did not target any of these common errors in their new web-based grammar checker.

In addition, these two grammar checkers cannot deal with two types of errors, sentence fragments and run-on sentences. The reason Microsoft ESL Assistant cannot deal with these two types of errors is not clear because the Microsoft checker uses both a rule-based approach and a statistical approach. A rule-based approach might help to resolve these two common ESL errors.

Nevertheless, the failure in detecting these two types of learner errors highlights one of the major weaknesses of the statistical grammar checker. The ngram-based statistical grammar checker is good at detecting errors in the “local” domains or “narrow” domains because the language models used by a statistic-based grammar checker are bigrams and trigrams. Therefore, it can thus find many errors based on inappropriate word combinations. Nevertheless, the statistical grammar checker has great difficulty in detecting errors that cannot be inferred based on word combinations. For instance, the errors like fragments and run-on sentences cannot be detected by a statistical checker because these errors are hard to detect solely based on the information of word combination. Similarly, errors like pronoun errors and tense errors cannot be detected effectively by the statistical grammar checker. To sum up, the statistical grammar checker will fail to capture errors if the errors are not word combination problems or they involve problems of non-adjacent word strings or conflicts across different clause boundaries.

When these two grammar checkers are examined in terms of precision and recall, Microsoft ESL Assistant evidently has high precision rate in several error categories (Missing or Extra Article, Subject-Verb Agreement, and Confused Words) but overall its average precision rate is only about 50%. In addition, ESL Assistant has much lower recall rate (only about 30%). On the other hand, the NTNU statistical grammar checker clearly has higher average recall rate (72%), but its precision rate (61%) is similar to Microsoft system because it often generated false alarms.

#### **4.1 Possible Directions for Improvement**

Even the state-of-the art grammar checkers like ETS Criterion and Vantage MyAccess have some limitations in dealing with ESL learners’ errors. It was found that they could not deal with errors like word order, tenses, and collocations (*cf.* Chen, 2006, 2009). It is clear that these two web-based grammar checkers have much room for improvement. The coverage and capacities of Microsoft ESL Assistant are still quite limited. This could be the major problem for ESL users around the world who use this service via Internet. Many types of ESL learner errors (spelling, subject-verb agreement) are not treated in the current version. The recent incorporation of the Bing search engine into the system is successful, and this change has made the system more user-friendly. Users now can quickly check their usage with the help of Bing. Microsoft has a very strong research team behind the system, and it should not be too difficult to deal with some common errors like spelling and subject-verb agreement. It is expected that the new version of Microsoft ESL Assistant would significantly enlarge their coverage.

Although the Microsoft Bing search engine can help users in some cases, the help it can provide is still limited. For some errors, Bing engine can search the native corpus and find

similar expressions and show them to learners. Making correct recommendations for ESL learners, however, is not an easy task in many cases. For example, it would be difficult if the errors are collocation errors. If a grammar checker needs to provide automatic feedback on a sentence like “I would like to increase my life,” it would be very difficult since there are so many English verbs that collocate with the noun “life”. Even if the system tries to find all the possible verbs, right suggestions cannot be provided easily. It seems obvious that Microsoft ESL Assistant might also need to improve its feedback mechanisms and provide more specific suggestions.

For the NTNU grammar checker, there are at least four major directions for improvement. First, the number of false alarms should be reduced. Although the NTNU grammar checker can be effective in detecting various learner errors, the sensitivity to learner errors can cause some false alarms. This noise can be rather annoying for ESL learners since they might not have the ability to judge if these messages are false alarms or real errors. Some possible ways of improving the ngram-based statistical grammar checker would involve a large reference corpus and perhaps the proper use of a POS-tagged reference corpus. The new data set, Web 1T 5-gram Version 1 contributed by Google Inc., contains English word n-grams and their observed frequency counts. The length of the n-grams ranges from unigrams (single words) to five-grams. The new huge corpus will be useful for statistical language modeling. The NTNU team is now developing another ngram-based checker based on the combination of Google data and BNC data (<http://140.122.83.245/gchecker/index2.html>), and we hope this new version with a much larger corpus will reduce some false alarms. Incorporating information from tagged corpora would be another way of improving the statistic-based checker. A better statistical checker might use the tag information to detect errors. Similar to raw text data, a part of speech (POS)-annotated corpus can be used to build a list of POS tag sequences. Some sequences will be very common (for example, determiner, adjective, noun as in “the old man”), others will probably not occur at all (for example, determiner, determiner, adjective). The powerful tagger can help us to tag large text files. The clusters of various POS tags can be further extracted and used to help to identify learners’ errors. The tagged corpus might also help to provide better grammatical explanations.

Second, the feedback mechanism can be further improved. The external link to the Google search engine can only solve some problems like word forms and misspellings. Google can only provide correct examples in some cases. If a student writes “...prepare the exam,” the student can easily find the correct answer “prepare for the exam” from Google. Nevertheless, Google cannot always find useful information based on the word strings provided by ESL learners. In addition, learners might need some metalinguistic feedback. It seems that a more robust feedback mechanism should be built.

Third, the highlighting techniques were not accurate enough. Sometimes the NTNU



grammar checker highlighted the adjacent words but not the keyword. Inaccurate highlighting can confuse ESL learners if they do not receive some explanations about using the grammar checker. A possible way of resolving this problem is to show the types of errors made by learners. If the checker can also indicate the types of errors that learners make, then the information can give learners more direct help.

Finally, as mentioned earlier, the limitation of statistical grammar checker is obvious. The NTNU checker would need to integrate the strength of the rule-based checker. The strength of a rule-based checker is that it can allow programmers to specify what needs to be found and what needs to be replaced. In the future, when a tagger or a chunker is integrated with the existing system, more grammar rules based on POS tags and structural relations can be added in to deal with more grammar problems like run-on sentences, fragments, conjunction errors, and tense errors. If the two different grammar checkers can be integrated into one system, then the integrated system should be able to detect more learner errors. (*cf.* Chen, 2006, 2009). If these four problems can be solved, the NTNU grammar checker can be more useful for ESL learners.

## **4.2 Limitations and Implications**

This study is a preliminary study on assessing the performances of two new web-based grammar checkers. Several limitations should be noted. There were only 10 types of errors tested. More types of errors from different L1 backgrounds should be included in future tests. In addition, it would be also interesting to compare human error corrections and machine corrections in future research. The strengths and limitations of machine feedback can be further identified.

It is indeed very challenging to develop a robust grammar checker. Researchers have identified some possible reasons. First, the errors made by second/foreign language learners are complex and diverse. There are so many different types of errors in learners' writing. A short sentence written by ESL learners might contain several errors. Because of the complexity, it is not easy to provide satisfactory feedback on these grammar errors made by learners. Second, the NLP technologies which are used to support these grammar checking tools are not in a mature stage. English parsers are still not perfect in detecting various errors and computers still cannot understand the meaning that learners try to convey. Many ESL errors cannot be detected, and appropriate suggestions for corrections cannot be provided.

For developers and researchers of Computer-Assisted Language Learning systems, despite the richer technology and corpus resources available now, the task for developing a robust grammar checker remains arduous and complicated. As indicated earlier, feedback plays a central role in the second language writing process. If specific and clear feedback messages can be provided to learners in the writing process, it is very likely learners can

incorporate the feedback to further improve their writing performance. A robust grammar checker can significantly reduce writing teachers' load and enhance students' writing performance. More research effort should be made to develop a robust grammar checker that can provide specific and clear feedback to ESL writers.

### Acknowledgments

This research was supported in part by National Science Council under Grant 96-2411-H-003-042-MY2.

### Reference

- Attali, Y. (2004). *Exploring the feedback and revision features of Criterion*. Paper presented at the National Council on Measurement in Education (NCME) in San Diego, CA.
- Atwell, E. & Elliott, S. (1987). Dealing with Ill-formed English Text. In: R. Garside, G. Leech and G. Sampson (Eds.), *The Computational Analysis of English: A Corpus-based Approach*. London: Longman.
- Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion online writing service. *AI Magazine*, 25(3), 27-36.
- Chen, H. J. (2006). Examining the Scoring Mechanism and Feedback Quality of My Access. *Proceedings of Tamkang University Conference on Second Language Writing*.
- Chen, H. J. (2009). Developing Statistic-based and Rule-based Grammar Checkers for Chinese ESL Learners. *Proceedings of the 2009 LITC International Conference on English Language Teaching and Testing*. Taipei: Taiwan.
- Chiu, T. L. (2008). An Evaluation of Whitesmoke: A Grammar-checking and Text-enrichment Software. *Proceedings of the Seventeenth ETA/ROC International Symposium on English Teaching*. Taipei: Taiwan.
- Chodorow, M. & Leacock, C. (2000). *An unsupervised method for detecting grammatical errors*. Paper presented at the 1st Annual Meeting of the North American Chapter of the Association for Computational Linguistics.
- Elliott, S. M. (2001). *IntelliMetric: From here to validity*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.
- Elliott, S. M., & Mikulas, C. (2004). *The impact of MY Access! use on student writing performance: A technology overview and four studies*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.
- Ferris, D. (1999). The Case for Grammar Correction in L2 Writing Classes: A Response to Truscott. *Journal of Second Language Writing*, 8(1), 1-11.
- Ferris, D. (2003). *Treatment of Error in Second Language Student Writing*. Ann Arbor, MI: University of Michigan Press.

- Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In Hyland, K. & Hyland, F. (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81-104). Cambridge: Cambridge University Press.
- Goldstein, L. (2006). Feedback and revision in second language writing: Contextual, teacher, and student variables. In Hyland, K. & Hyland, F. (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 185-205). Cambridge: Cambridge University Press.
- Guenette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing*, 16, 40-53.
- Han, N.-R. Chodorow, M., & Leacock, C. (2006). Detecting errors in English article usage by nonnative speakers. *Natural Language Engineering*, 12(2), 115-129.
- Higgins, D., Burstein, J., & Attali, Y. (2006). Identifying off-topic student essays without topic-specific training data. *Natural Language Engineering*, 12(2), 145-159.
- Holland, M. V., Kaplan, J.D., & Sama, M.R., eds. (1995). *Intelligent Language Tutors: Theory Shaping Technology*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Jensen, K. (1993). PEG: the PLNLP English Grammar, in Jensen K., Heidorn G.E., & Richardson S.D., (Eds.), *Natural Language Processing: The PLNLP Approach*, Kluwer Academic Publishers, Boston, 29-45.
- Liou, H. C. (1991) Development of an English grammar checker: A progress report. *CALICO Journal*, 9(1), 57-70.
- Liou, H. C. (1992). An automatic text-analysis project for EFL writing revision. *System*, 20(4), 481-492.
- Liou, H. C. (1993). Integrating text-analysis programs into classroom writing revision. *CAELL Journal*, 4(1), 21-27.
- Moré, J. (2006). A Grammar Checker Based on Web Searching. Digithum [online article]. Retrieved October 14, 2009, from <http://www.uoc.edu/digithum/8/dt/eng/more.pdf>.
- Naber, D. (2003). *A Rule-Based Style and Grammar Checker*. Unpublished doctoral dissertation, University of Bielefeld, Germany.
- Park, J. C., Palmer, M., & Washburn, G. (1997). An English grammar checker as a writing aid for students of English as a second language. *Proceedings of the Conference of Applied Natural Language Processing (ANLP)*. Washington, DC.
- Sjöbergh, J. (2005). Chunking: an unsupervised method to find errors in text. *Proceedings of the 15th Nordic Conference of Computational Linguistics, NODALIDA 2005*.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46, 327-369.
- Truscott, J. (1999). The case for “The case against grammar correction in L2 writing classes”: A response to Ferris. *Journal of Second Language Writing*, 8(2), 111-122.
- Truscott, J. (2007). The effect of error correction on learners’ ability to write accurately. *Journal of Second Language Writing*, 16, 255-272.

Wu, S.-H., Su, C.-Y., Jiang, T.-J., & Hsu, W.-L. (2006). An Evaluation of Adopting Language Model as the Checker of Preposition Usage. *Proceedings from ROCLING 2006*.