MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

The English Writerś Assistant: A system for helping non-native speakers write English text

Andrew Golding, Emmanuel Roche, Yves Schabes

TR96-12a December 1997

Abstract

Writing text in English presents a challenge to non-native speakers because of the difficulties of mastering English vocabulary, grammar, and usage. The English Writerś Assistant is a software system that helps non-native speakers write correct English sentences. The system detects grammatical errors and suggests corrections. It is based on a novel statistical approach to language processing, and is much more powerful than commercial grammar checkers. A prototype version of the English Writerś Assistant has been built. The system runs under Windows95 and Unix.

MERL – A MITSUBISHI ELECTRIC RESEARCH LABORATORY
http://www.merl.com

The English Writer's Assistant: A system for helping non-native speakers write English text

Andrew Golding, Emmanuel Roche, and Yves Schabes

TR-96–12a   May 1997

Abstract

Writing text in English presents a challenge to non-native speakers because of the difficulties of mastering English vocabulary, grammar, and usage. The English Writer's Assistant is a software system that helps non-native speakers write correct English sentences. The system detects grammatical errors and suggests corrections. It is based on a novel statistical approach to language processing, and is much more powerful than commercial grammar checkers.

The system can be used in such applications as E-mail composition or general text composition. It can also be regarded as an educational system, in that users learn to improve their English by watching the corrections the system makes.

A prototype version of the English Writer's Assistant has been built. The system runs under Windows95 and Unix.

1. First printing, TR96–12, January 1996
2. Second printing, TR96–12a, May 1997

1 Introduction

Writing text in English presents a challenge to non-native speakers because of the difficulties of mastering English vocabulary, grammar, and usage. Although most word-processor programs provide some kind of automatic grammar checking, these programs are inappropriate for non-native speakers for two reasons. First, these programs are designed specifically for native speakers, and thus they only address the mistakes that native speakers make. Non-native speakers make a different set of errors, strongly dependent on their own native language. Second, these programs are usually not very efficient, or they report trivial style errors (such as the use of passive voice) for which no consensus exists.

We have designed a system, The English Writer's Assistant (EWA), specifically for detecting and correcting the grammatical errors of non-native speakers, and in particular Japanese speakers. The system can be used in such applications as E-mail composition or general text composition. In addition to its grammar-checking facility, the system also provides a suite of related tools for helping the user write English text. An additional benefit of the system it that it can be regarded as an educational system, in that users learn to improve their English by watching the corrections the system makes.

The system is based on a novel statistical approach to language processing. The present system, although in a prototype stage, is already much more powerful than commercial systems. The current version of the system runs under Windows95 and Unix.

The main features of the software are as follows:

- The system embodies a new statistical approach to language processing.
- It detects grammatical errors in a context-sensitive manner.
- It outperforms commercial systems.
- It can be used with word processors and E-mail systems.
- It runs under Unix and Windows95.

The prototype implementation is embedded in a word processor. The main window of the program is shown in Figure 1. It consists of a large frame in which text can be edited as in most editors. It also has various controls, menus, and buttons, through which all the standard text manipulation operations can be performed. In addition, it provides access to a set of tools specific to the correct use of English. These tools, detailed below, include a spelling checker, a grammar checker, morphological tools, and a specialized dictionary access command.

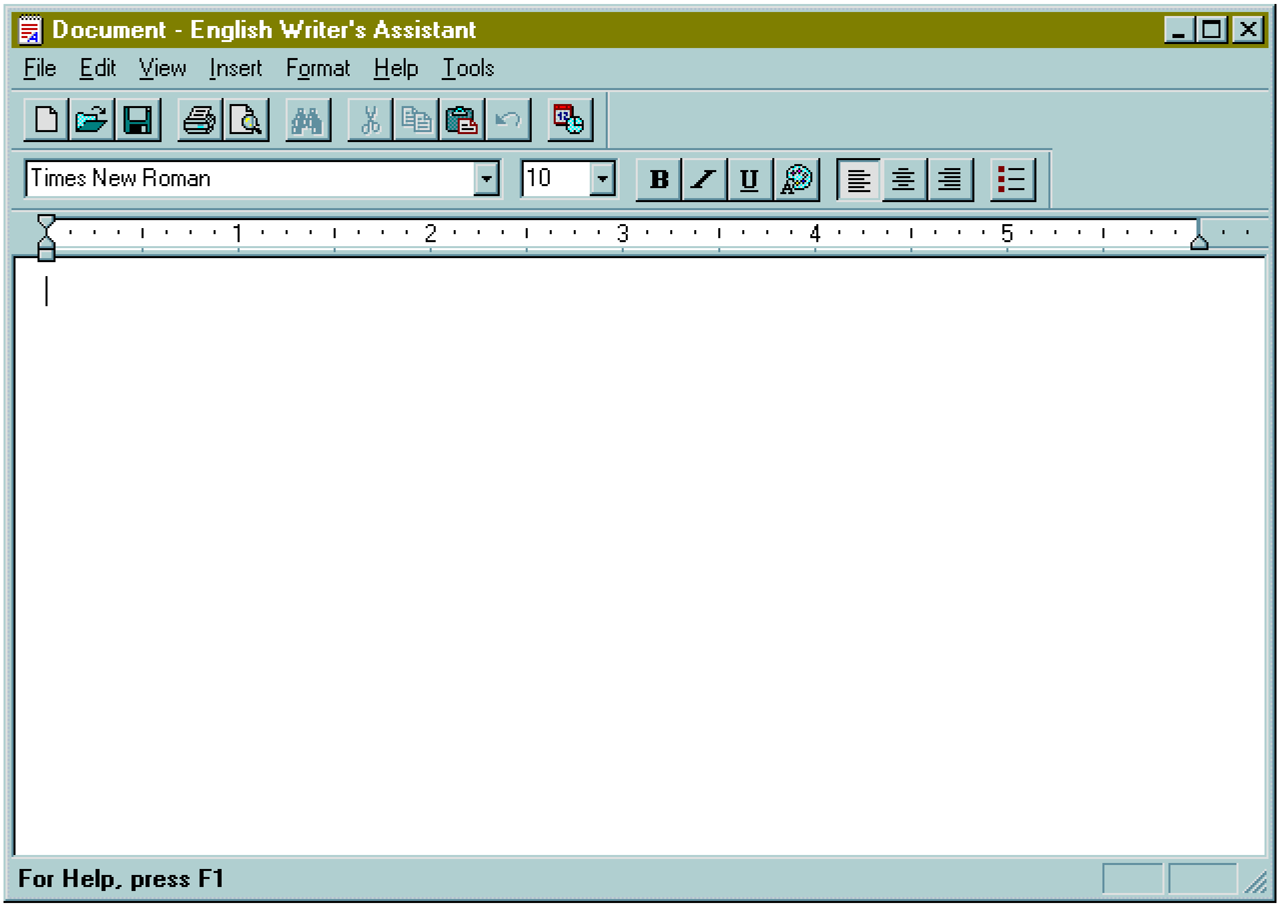
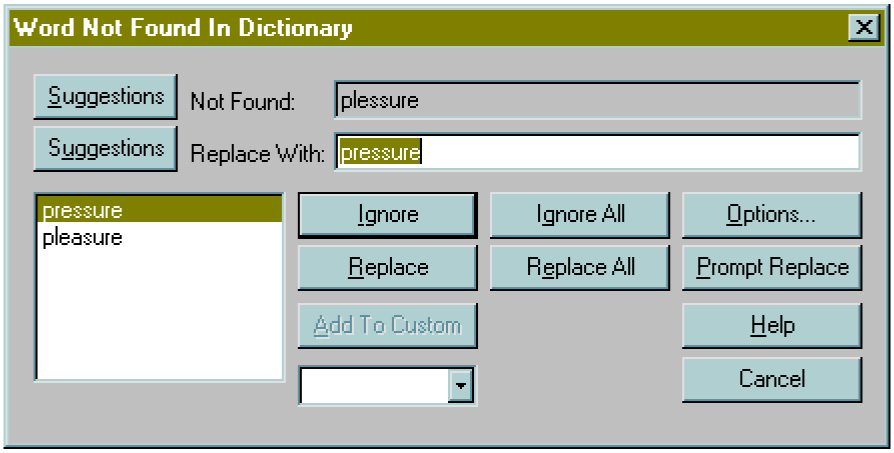

Document - English Writer's Assistant
File Edit View Insert Format Help Tools
Times New Roman
10
B I U
1 2 3 4 5
For Help, press F1

Figure 1: EWA Main Window.



Not Found: plessure

Replace With: pressure

pressure
pleasure

Figure 2: Spelling Window.

2 Applications

In this section, we describe applications for which tools for writing correct English are of particular interest.

Obviously, a word processor, whether it is used to write business letters, technical descriptions, internal notes, or other documents, can benefit greatly from a tool that helps the writer produce better quality text. This is of particular importance for global companies which have a great deal of business overseas and for whom communication with international partners is an important bottleneck. Improvements in the quality of text produced naturally help promote an image of seriousness and openness, and in some cases can help avoid misunderstandings that might otherwise be extremely costly.

A second very natural application consists of grammar-checking tools within E-mail applications. In fact, the volume of international business transactions going through the Internet is currently growing at an exponential rate, and this growth is likely to continue in the near future. Even if readers are more tolerant of errors in E-mail than in conventional written letters, clarity in E-mail messages is still crucial. Indeed, E-mail transactions are carried out very much like face-to-face conversations, and it can take a long time for each speaker to be aware of any misunderstanding. This is especially true if the channel is noisy, for example if both speakers produce poor quality English text.

The tools used in our system are also very well adapted to English teaching. In fact, since the system is especially tuned to correct classical non-native speaker mistakes, a course focusing on any of these mistakes could take great advantage of the English Writer's Assistant. The teacher could, for instance, propose an exercise for a typical non-native speaker problem, and the computer would provide the student with instantaneous feedback, pointing out problems the student overlooked. The student can then further explore the domain, at his/her own pace, making sure that the problem is properly understood.

3 Description of the System

The software is based on a novel statistical approach to language processing. Because of this novel approach, it outperforms current commercial grammar-checking systems. Another consequence of the approach is that the system can easily be extended to mistakes specific to a user or to a group of users.

The core of the system is a grammar checker that, given an input sentence, tries to find English mistakes and propose corrections for them. Sometimes, if the original sentence is unclear, the best the system can do is point out that there is a problem and leave it to the user to find a better way of expressing his/her thought.

Let us now describe the classes of grammatical errors that the current system corrects:

- Auxiliary verb sequences, such as:

  (1) He should have consider this fact.
  (2) ⇒ He should have considered this fact.

  This is a very typical error. The length and complexity of some auxiliary verb sequences make it difficult, even for experienced speakers, to use the correct tense of the verb.

- Missing determiners, such as:

  (3) He is good writer.
  (4) ⇒ He is a good writer.

(5) Knowledge is stored in human brain.
(6) ⇒ Knowledge is stored in the human brain.

This problem is common among native speakers of Japanese and other Asian languages, for which the use (or absence) of an article appears to be an extraordinary puzzle. In fact, this problem is a challenge for English grammarians since there is no clear set of rules to apply, but rather many word-specific cases to remember.

- Lack of agreement between the determiner and noun, as in:

  (7) He has many useful tool.
  (8) ⇒ He has many useful tools.

  This kind of mistake is very common among both native and non-native speakers. Long noun phrases and fast typing are the usual causes of this problem, rather than a lack of knowledge of English syntax.

- Easily confused words, such as piece/peace, to/too/two, except/accept, then/than, “may be” and “maybe”, etc.:

  (9) Sending E-mail too the USA is a peace of cake.
  (10) ⇒ Sending E-mail to the USA is a piece of cake.

  These lists of easily confused words are well known by teachers of English as a second language. These confused words are often pronounced in a similar way, but used in completely different contexts. Sometimes, as in the case of *to* versus *too*, they are grammatical words that have very different functions; in other cases, as in *desert* versus *dessert*, they have a similar syntactic function (both nouns), but mean completely different things. In both cases, it is crucial to avoid the mistake, as it makes understanding the text quite difficult.

- Incorrect spelling of verbs, nouns, and adjectives, such as “drived”, “childs”, and “gooder”:

  (11) He drived to New York yesterday.
  (12) ⇒ He drove to New York yesterday.

The morphology of English, as for many other languages, is extremely irregular; as any student of English knows, finding the right inflection of a verb, noun, or adjective can be a difficult task. It is natural to try to inflect an irregular verb, such as *to drive*, as if it were regular. This leads to an incorrect form, such as *drived*, whereas the correct inflection, *drove*, is entirely specific to the verb. The system presented here detects such mistakes and proposes the correct inflection.

- Incorrect preposition usage, such as:

  (13) She is in home.
  (14) ⇒ She is at home.

  (15) He is in pressure to perform.
  (16) ⇒ He is under pressure to perform.

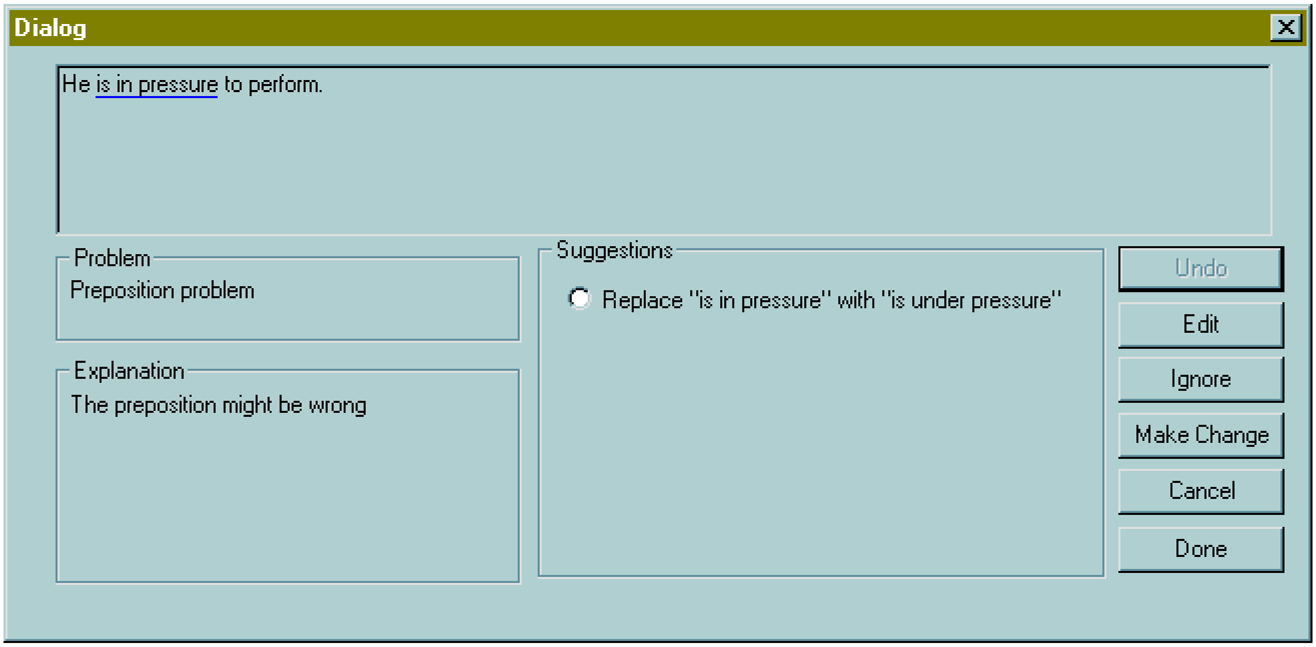
  Knowing which preposition to use with a particular verb is one of the worst nightmares of any student of English whether he/she is just a beginner or has many years of experience. In fact, this problem is also encountered by native speakers. The reason prepositions are difficult is that each verb has its own idiosyncratic use of prepositions; no simple set of rules can help the non-native speaker predict them accurately. The speaker is left with the unpleasant option of simply memorizing the usage for all verbs. This is extremely difficult, as the number of verb/preposition pairs is on the order of tens of thousands.

- Subject-verb agreement:

  (17) The course taught by those teachers were tough.
  (18) ⇒ The course taught by those teachers was tough.

  The sentence above is a typical example in which it is easy to violate the rule that the subject and verb must agree in number. This problem arises most often when the subject and verb are far apart within the sentence, separated by relative clauses or adverbial phrases.

These classes account for tens of thousands of actual grammatical errors in text. Figure 3 shows an instance of the system detecting a mistake. The



He is in pressure to perform.

Problem
Preposition problem

Explanation
The preposition might be wrong

Suggestions
○ Replace "is in pressure" with "is under pressure"

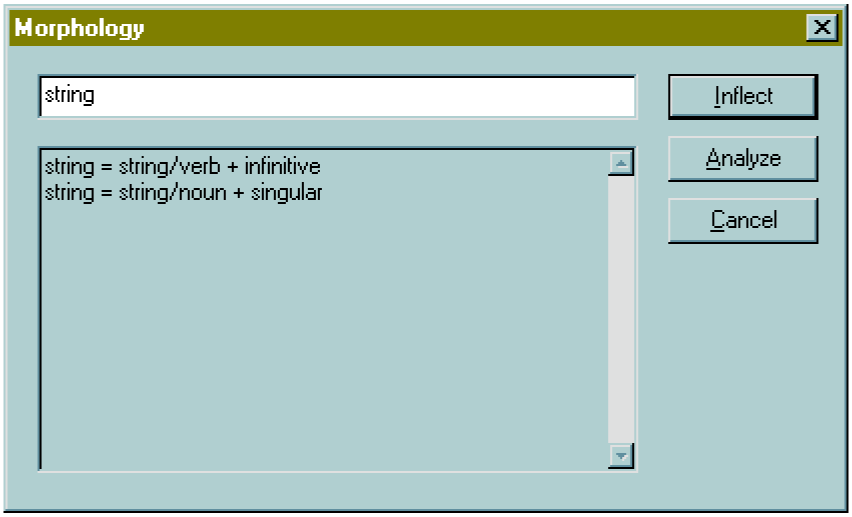
Figure 3: Grammar Window.

system displays the sentence, and underlines the sequence of words that appear to be incorrect. It gives a concise description of the problem and an explanation. It also proposes a correction. The user can accept the proposed correction, edit the sentence himself/herself, or ignore the problem. Once this is done, the system goes on to the next problem (which might be in the same sentence or in another one).

In addition to correcting the grammar, the system provides tools for helping the user write English text. Such tools provide the following functionality:

- Analyze a word morphologically.
- Generate the inflections (past tense, plural, etc.) of a word.
- Look up a word in a dictionary and show only the entries relevant to the context in which the user used the word.
- Explain the corrections that the system suggests.

Figure 4 shows an example of the morphological analysis of a word. Figure 5 shows an example of generating all possible inflections of a word.

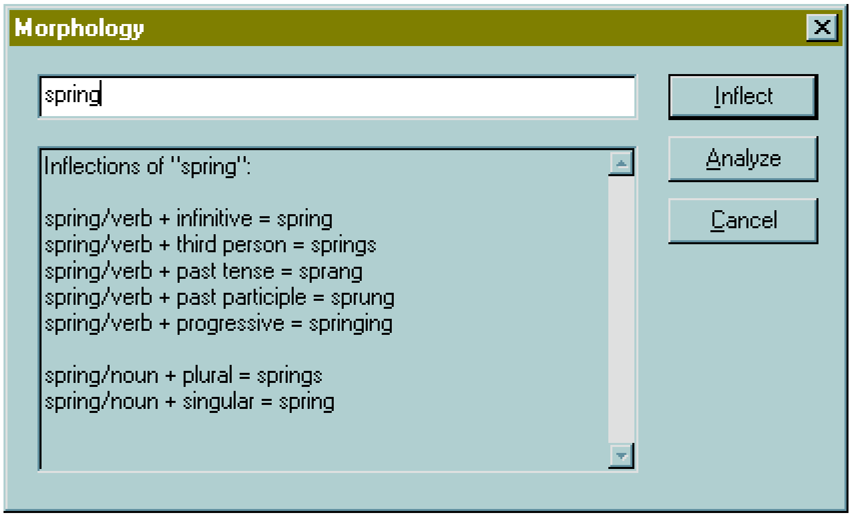


string = string/verb + infinitive
string = string/noun + singular

Figure 4: Analyze Window.

The system is also able to look up any word that appears in the text. It shows the definition of the selected word according to the context in which the word appears. In other words, if the word is ambiguous, such as *left*, which can be a verb, noun, adjective, or adverb, the system first gives the definition of the word as it is used in the current sentence. For instance, in the sentence is *He made a left turn*, the system, when asked to look up *left*, will first give the definition of *left* as an adjective. At the same time, another window displays the other definitions of the word. Figure 6 shows what the system displays when the word *left* has been looked up in the preceding example. It is also possible to look up a word out of context as demonstrated in Figure 7.

The system can be tuned to a particular user and to his/her proficiency in English. For example, a user that already has a thorough knowledge of English might need a correction such as *too* versus *to* only if the system is absolutely certain that the input sentence is incorrect; whereas a user with less experience might feel more comfortable with a system that points out a problem as soon as there is any doubt about correctness. Figure 8 shows a control window that allows the user to set the sensitivity of the correction



spring

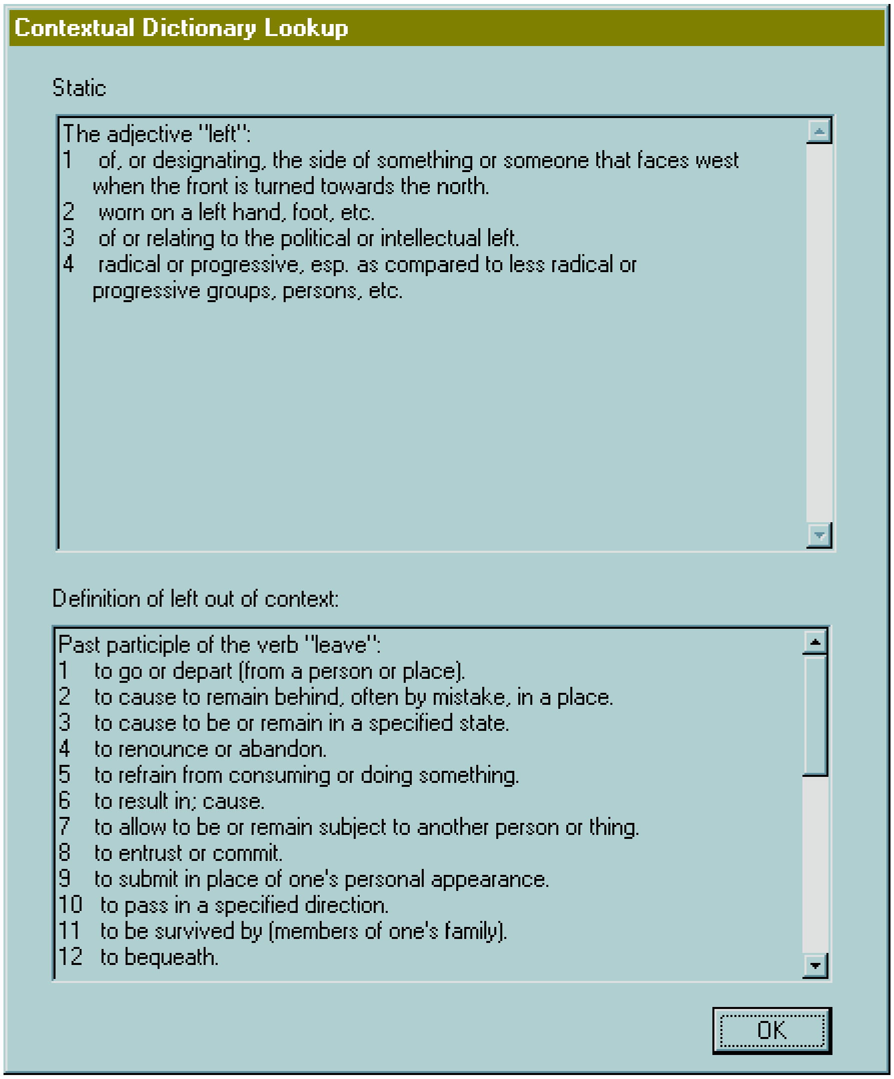
```
Inflections of "spring":

spring/verb + infinitive = spring
spring/verb + third person = springs
spring/verb + past tense = sprang
spring/verb + past participle = sprung
spring/verb + progressive = springing

spring/noun + plural = springs
spring/noun + singular = spring
```

Figure 5: Inflection Window.

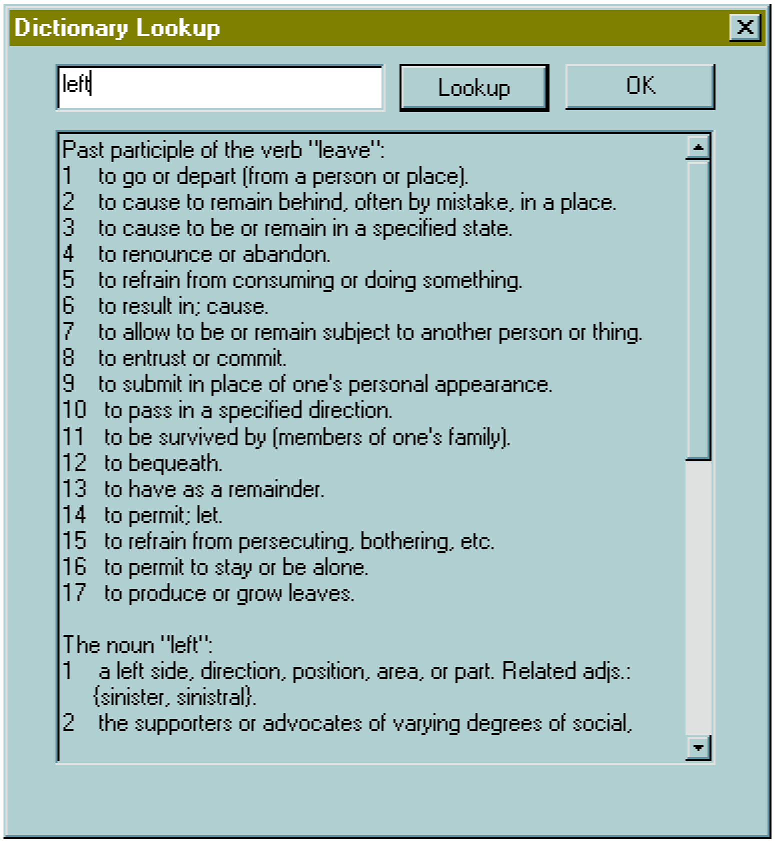
such that it fits his/her personal needs.


Contextual Dictionary Lookup
Static
Definition of left out of context:
OK

The adjective "left":
1 of, or designating, the side of something or someone that faces west when the front is turned towards the north.
2 worn on a left hand, foot, etc.
3 of or relating to the political or intellectual left.
4 radical or progressive, esp. as compared to less radical or progressive groups, persons, etc.

Past participle of the verb "leave":
1 to go or depart (from a person or place).
2 to cause to remain behind, often by mistake, in a place.
3 to cause to be or remain in a specified state.
4 to renounce or abandon.
5 to refrain from consuming or doing something.
6 to result in; cause.
7 to allow to be or remain subject to another person or thing.
8 to entrust or commit.
9 to submit in place of one's personal appearance.
10 to pass in a specified direction.
11 to be survived by (members of one's family).
12 to bequeath.

Figure 6: Context-sensitive dictionary lookup.



Past participle of the verb "leave":
1 to go or depart (from a person or place).
2 to cause to remain behind, often by mistake, in a place.
3 to cause to be or remain in a specified state.
4 to renounce or abandon.
5 to refrain from consuming or doing something.
6 to result in; cause.
7 to allow to be or remain subject to another person or thing.
8 to entrust or commit.
9 to submit in place of one's personal appearance.
10 to pass in a specified direction.
11 to be survived by (members of one's family).
12 to bequeath.
13 to have as a remainder.
14 to permit; let.
15 to refrain from persecuting, bothering, etc.
16 to permit to stay or be alone.
17 to produce or grow leaves.

The noun "left":
1 a left side, direction, position, area, or part. Related adjs.: {sinister, sinistral}.
2 the supporters or advocates of varying degrees of social,

Figure 7: Simple dictionary lookup.

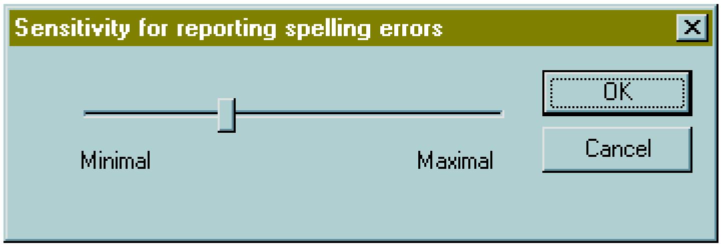


Figure 8: Sensitivity Window.