



State Of The Art Word Processing Tools

GRAMMATIK^{T.M.}

BEYOND SPELLING CHECKING

Software
for
CP/M[®]
IBM Personal Computer

Aspen Software Co.

P.O. Box 339 Tijeras, NM 87059 - (505) 281-1634



Grammatik

Beyond Spelling Checking

For

CP/M

IBM Personal Computer

1.0	Introduction	3
1.1	Backing up distribution disk	4
1.2	Organizing files	5
2.0	Grammatik System Overview	6
2.1	GMK - Grammatik	6
2.2	PHRASES.GMK - phrase dictionary	6
2.3	SEXIST.GMK - sexist terms	6
2.4	PROFILE - word use profile	6
2.5	SORTDICT - sort utility	6
3.0	Analyzing a Document	7
4.0	Specific Operating Procedures	11
4.1	GMK	11
4.1.1	Starting up the first time	11
4.1.2	Interactive command menu	12
4.1.3	Automatic Configuration	16
4.2	PROFILE	18
4.3	SORTDICT	18

APPENDICES

A --	Error messages	19
B --	Dictionary structure	21
C --	Special words recognized by Grammatik ...	22
D --	Bibliography	23
INDEX	24

***** IMPORTANT! *****

Please follow instructions in section 4.1.1
when using Grammatik for the first time!

Copyright (c) 1981,1982 Aspen Software Company
All rights reserved
Reproduction by any method prohibited

Grammatik, Proofreader, Proof-Edit trademarks of Aspen Software Co.

Grammatik User's Manual
CP/M

IBM Personal Computer

1.0 Introduction

Grammatik(tm) is a document checking system that includes programs and data bases used to analyze writing style at both the word and sentence level, as well as performing checks for common typographical errors. Grammatik and a spelling checker such as Aspen Software's Proofreader provide the writer with a powerful automated document proofreading system.

Grammatik provides two areas of checking capability: writing style and typographical errors. By "style" is meant a writer's particular choices for words and sentence forms. Although what is good or bad style is a subjective judgement, particularly for word choice, there are some standards that experts agree lead to good style. While a document that conforms to the rules of "good style" is not guaranteed to be coherent and readable, a document that fails to follow the rules is more likely to be confusing and difficult to read. Style analysis performed by Grammatik includes checks for specific phrases commonly recognized as being poor or wordy usage, as well as compilation of statistics about word and sentence length.

Besides the style analysis, Grammatik checks for other typographical errors commonly found in text generated on computers. Probably the most common errors Grammatik will detect are doubled words and punctuation marks such as "the the" or ",," and inconsistent capitalization ("STicky shift key"). It also checks for capitalization of the first word of a sentence, balanced quotation marks and parentheses, and placement of punctuation inside quotation marks.

As a separate procedure, a document can be profiled for word usage. Each different word in the document is listed indicating how many times it was used. Besides being interesting, this information is useful for discovering spelling errors and overworked vocabulary.

Grammatik has been designed to work with most CP/M and IBM Personal Computer word processors. The commands and operating procedures used by Grammatik are nearly identical for all systems.

1.1 - Backing up distribution package

The Grammatik distribution disk contains several programs and dictionary data bases used to analyze documents. Before using any of the Grammatik programs, you should make a copy of all the files on the distribution disk.

Once you have made backup copies, please read the manual before using the program. If you can't wait, follow the directions in section 4.1.1!

1.1.1 - Backing up: CP/M

On CP/M, Grammatik is distributed as a standard CP/M data disk. Before attempting to use Grammatik, you should make a backup of all the files contained on the distribution disk. The method used to make the backup will vary from system to system. Either two physical disk drives or a CP/M system that supports "virtual" drives or one drive copies on a one drive system will be required to make a backup.

The easiest way to backup Grammatik is to first create a fresh copy of a CP/M system disk. Format a new disk if required, then use the SYSGEN utility to move a copy of CP/M to the new disk. More details about SYSGEN are given in the "AN INTRODUCTION TO CP/M FEATURES AND FACILITIES" manual. Some versions of CP/M may use a different procedure to create a new system disk.

After you have created the new system disk, you should also put PIP and other day to day utilities on that disk. Assume that the newly created system disk is mounted in drive B:. You can copy old files to the new disk with PIP. For example:

```
PIP B:=A:PIP.COM
```

After you have copied at least PIP to the new disk, remove the old CP/M disk, and put the newly created system disk in A:. After you reset ("cold boot") the system with that new disk in A:, mount the Grammatik distribution disk in drive B: and use PIP to transfer the files on the Grammatik distribution disk to the new disk:

```
PIP A:=B:*.*
```

This backup procedure will apply to the vast majority of CP/M systems. If your system has a different backup procedure, or has limited disk space available, you may have to use different procedures.

1.1.2 - Backing up: IBM PC and MS-DOS

The procedure for backing up Grammatik on MS-DOS is very similar to the procedure used for CP/M. Read the CP/M description first if you haven't already. Create a fresh system disk with no extra utilities or files. Put that new disk in A:, and use the COPY (instead of the equivalent CP/M PIP) command to copy all the files from the Grammatik distribution disk to the new disk: "COPY B:*. * A:". That disk is now ready to use.

1.2 - Organizing files

Depending on your disk drive configuration, you will find that life will be easier if you put some thought into how you organize your files. The part of Grammatik required for day to day operation requires about 50,000 bytes of free disk space. The PROFILE utility requires about another 10,000 to 15,000 bytes.

For systems with disks with a capacity of 250,000 bytes or more, the best organization is probably to keep your word processor, your spelling checker such as Aspen Software's Proofreader, Grammatik, and useful day to day utilities (such as PIP and STAT on CP/M or CHKDSK on MS-DOS) on a single disk used from drive A:. Then, use drive B: for disks containing your documents.

If your disks do not have enough room for both Grammatik and your word processor, then it is probably best to keep your word processor on one disk used in drive A:, and the Grammatik files on a second disk used in drive A:. You can still keep your documents on disks used in drive B:.

2.0 Grammatik System Overview

Each of the files included with the Grammatik distribution disk is summarized in the following sections. An overview of the document analysis process is contained in section 3.0, and specific details for using the programs contained in the package are given in section 4.0. Program files in CP/M use an extension of ".COM". IBM Personal Computer DOS program files use a ".EXE" extension. Because the ".COM" or ".EXE" is not normally used when running a program, the descriptions in this manual will not use the ".COM" or ".EXE". The operation of Grammatik is almost identical on both CP/M and the IBM Personal Computer (or other MS-DOS 8086 based systems).

2.1 GMK - Grammatik

The main document analysis program is called GMK.COM (or GMK.EXE). This program reads phrase dictionaries and checks the document.

2.2 PHRASES.GMK - phrase dictionary

PHRASES.GMK is the main phrase dictionary used by Grammatik. The standard version includes over 500 phrases and words used in analyzing the source document. Included in the dictionary are the words or phrases, error categories, and suggestions for alternatives. This dictionary may be modified by the user as needed for particular applications.

2.3 SEXIST.GMK - sexist terms

SEXIST.GMK is a dictionary of about 100 words that are gender specific and possibly sexist. For those writers desiring to remove such terms from their documents, this dictionary may be useful. It can be used alone, with PHRASES.GMK, or with other user supplied dictionaries.

2.4 PROFILE - word use profiler

This program is used to profile word usage in a document. Each different word in the document is listed with the total number of times it was used indicated.

2.5 SORTDICT - sort utility

SORTDICT is a utility that can be used to sort phrase dictionaries. It can also be used to sort other text files.

3.0 Analyzing a Document

This section contains an overview of the process used to analyze a document. Specific procedures for starting Grammatik and setting operation options are given in section 4.0. Grammatik is menu driven and easy to use. It is possible to begin checking documents almost immediately by following the instructions in Section 4.1.1. Before trying to use Grammatik, however, it is probably best to read over this section to get an idea of how the program operates.

The procedure for creating, analyzing, and revising a document consists of the following steps:

1. Create document with a word processor.
2. Correct the spelling errors. This step is important since Grammatik expects words to be spelled correctly.
3. Analyze document with Grammatik.
4. Revise document with word processor.
5. Repeat process until satisfied.

This section will describe in some detail the analysis procedure performed by Grammatik. After Grammatik has been started, it will read in a series of commands. Depending on the exact method used to start Grammatik, the commands may be read in from a file automatically, or they may be read in from an interactive command menu. Details are given in section 4. The commands will typically instruct Grammatik to read in a data base consisting of any number of dictionary files. These can optionally include the standard PHRASES.GMK, other dictionaries supplied by the user, the unknown word list generated by Proofreader, or no dictionaries at all.

After the dictionaries have been read, several options may be selected. These options include which errors to detect, where to display the error messages, and the format of error message output. Which dictionaries are used and which options are selected may be configured automatically, or may be set interactively depending on the exact procedure used to start Grammatik. The details are described in section 4.1.

Once the dictionaries have been read and the options selected, the source document is read and analyzed. Analysis takes place on word, phrase, and sentence levels. A word is a sequence of letters a to z (upper or lower case) separated from other words and symbols by characters that are not letters. Numbers and certain abbreviations are recognized as units by Grammatik. All other symbols are treated by Grammatik as individual characters. Grammatik can recognize any sequence of words or characters as a phrase if the sequence is contained in the phrase dictionary.

Grammatik recognizes certain punctuation marks. Periods, question marks, and exclamation marks are recognized as end of sentence marks. Periods used in numbers, "...", and recognized abbreviations are not counted as the end of a sentence. Double quotation marks " and the parenthesis pairs "()", "{}", and "{}" are recognized and counted to make sure they are balanced. The punctuation marks ",", and ";", and all words are checked for doubling like ".,," or "the

the". When the end of the sentence is detected, the next word is checked to be sure it begins with a capital letter. All words are checked to make sure that they are either all upper or lower case, or that only the first letter is capitalized (the word is assumed to be a proper name in that case). The single letter "I" is checked to be sure it is capitalized.

The text from the file is displayed on the screen as it is read. Undisplayable characters are displayed as an underscore ("_"). As words and phrases in the file are analyzed, Grammatik checks to see if they are contained in the phrase dictionary. If a word or phrase is found in the dictionary and is classified as an error, a message is displayed on the screen immediately following the line containing the error. For example:

```

+-----+
|
| : segment of source text. It has a number of
| : -----> a number of
| : * At sentence 23 - 'W' - Wordy phrase
| : * Suggestion: several, many, some
| : * Press: RETURN to continue
| : * or Q first to quit checking _
|
+-----+

```

The phrase, word, or punctuation mark in question is displayed after the '----->', "a number of" in the example. The previous line on the screen will contain the context of the item in question. The line following the '----->' gives the sentence number that the error was found in. This is the sentence number from the beginning of the document, and is an indication of the location of the error within the document. An error code, in the example 'W', is given, followed by an error message. In the example, "Wordy phrase" is the type of error detected. The errors detected by Grammatik are classified into several types, such as "wordy phrase" or "capitalization error". Detection of any specific error types can be optionally suppressed by the user.

The last line of the error message will usually give a suggestion for fixing the problem detected. In the example, the phrase "a number of" can often be replaced by the more concise words many, several, or some. It must be remembered that any of the errors detected may in fact be the correct or best choice for a given situation, and that the suggestion offered may not be the best alternative. Often an entire sentence or paragraph will require revision. The error message may also be ignored. Repeated flagging of a particular phrase in a single document indicates overuse.

Grammatik has several options for controlling how errors are indicated. Normally, the entire document is displayed on the screen, and errors listed when they are detected. Depending on the error display options set, Grammatik may pause after the error message. Pressing just the RETURN key (ENTER on some computers) causes error checking to continue. Pressing the "Q" key first, then RETURN causes checking of the current document to be aborted, and control is returned to the main command menu. The output to the screen may also be stopped at any time simply by pressing Control-S. Pressing Control-S again will allow scanning to continue. The error messages only optionally can be directed to the printer or to a disk file. The entire source file can be optionally copied to a new file with all the errors in the text marked with the symbol "?#x", where "x"

is the error type.

After checking has been completed, the file with the errors marked may be edited with the word processor to make required revisions. The global search command can be used to find all occurrences of "?#", which mark errors. The "?#x" should be deleted, and the error corrected if required. A listing of the errors and suggestions produced on the printer can be useful while correcting the marked error file.

At the end of checking Grammatik will print out a summary of statistics gathered during the analysis.

```
-----+-----
:                                     :
: Summary for EXAMPLE.TXT / Problems detected: 28 :
:                                     :
: # sent: 204 ; words: 3571 :
: avg sent len: 17.3 ; avg word len: 4.3 :
: # questions: 2 ; # imperatives: 1 :
: short sent(<14 wds):98 ; long sent(>30 wds):24 :
: longest 45 wds at #28 ; shortest 4 wds at #84 :
: to be's: 153 ; prepositions: 302 :
: User category totals: :
: #3: 29 #4: 9 :
:                                     :
: Enter: E or just RETURN to exit :
: or A to check another file: _ :
:                                     :
:-----+-----
```

The first line of the summary gives the file name and the number of problems detected by Grammatik. The next five lines give statistics about word and sentence length. The total number of sentences and words is shown on the first line. These totals indicate the overall document size. The next four lines indicate the average sentence and word length, the number of sentences ending in "?" or "!", and other information about sentence length.

Although these measures will not indicate whether the text is coherent and well organized, documents with many long sentences and polysyllabic words tend to have stylistic difficulties. If a document has a high technical content, or is intended to be an instruction manual, it is probably best to make the text easy to read. Lowering the average sentence and word length, as well as reducing the number of long sentences is one measure of increased readability.

The next line of the summary indicates usage of forms of the verb "to be" and common prepositions. These categories also give some measure of style. The totals for "to be's" and "prepositions" should be compared to the total number of sentences. A high ratio may indicate overuse of these word classes, and suggests a possibility for revision.

The final line lists total counts for user categories. The user may define up to seven different categories. Categories may consist of single words or phrases to count, or a group of words or phrases to count. For example, the pre-defined class of "to be's" is actually category #1. Similarly, prepositions

are actually category #2. The seven categories 3 to 9 are reserved for user definition. For example, it might be desired to count occurrences of "that" and "which". Category #3 could be assigned to "that", and #4 to "which". These two categories could be added to the dictionary as described in Appendix B.

After the summary has been displayed, the user is given the choice of exiting Grammatik, or checking another document. Entering an E followed by a RETURN, or simply RETURN will exit back to CP/M. If an A is entered, Grammatik will return to the main command menu. All statistics about the document will be cleared, but the dictionary will remain intact. A new document may then be analyzed without the need to re-read the dictionary.

While Grammatik does perform extensive analysis of writing style, the user must be cautioned that there are things that Grammatik will not do. Specifically, Grammatik does not know anything about the meaning of words. Anything that requires an understanding of meaning, such as subject-verb agreement, cannot be detected. Such software technology is still some years away. And, as mentioned before, many of the problems detected by Grammatik will be correct in some situations. However, if many problems are detected during text analysis it is a strong indication of a need to revise the document. After a document has been checked by Grammatik, it should still be carefully checked by a human for other possible errors.

4.0 Specific Operating Procedures

This section contains the operating instructions for the programs included in the Grammatik package.

4.1 GMK

GMK is the program used to analyze text. It may be invoked from the CP/M command level in one of two formats. After the CP/M "X>" prompt, enter one of the following command line formats:

GMK

or

GMK FILENAME.EXT

The first format, where just "GMK" is specified, invokes Grammatik and causes a command menu to be displayed. Commands for specifying the file to be checked, loading dictionaries, and controlling error listing format may then be entered from the keyboard. It is probably best to use that start up procedure while still unfamiliar with Grammatik's operation.

The second format, "GMK FILENAME.EXT", includes the name of the file to be checked on the command line. When this format is used, Grammatik will automatically look for a user supplied file called "CONFIG.GMK". That file will contain commands for automatically loading dictionaries and setting system options. This form is most convenient when using a standard set of dictionaries and options. Use of a configuration file is optional, and should NOT be attempted until a good understanding of Grammatik operation is achieved using the interactive command menu. More details about automatic configuration are included in section 4.1.3. The format of the command menu will be explained first, since it is necessary to understand it before generating configuration files.

If a switch in the form "/B" is included following the file name (e.g., "GMK MYFILE.TXT/B"), then Grammatik will automatically read in a dictionary using the file name with a ".BWD" extension (e.g., "MYFILE.BWD"). This switch will not normally be used, and is provided to allow an unknown word list to be passed from Aspen Software's spelling checker Proofreader directly to Grammatik.

WordStar users may start Grammatik from the "NO-FILE" command menu with the "R" run a program command. Specify "GMK FILENAME" in response to the "COMMAND?" prompt, and Grammatik will check the file. Control will be returned to WordStar after checking is complete.

4.1.1 -- *** Starting up the first time ***

To use Grammatik for the first time, the following procedure is suggested. Assume that the file called "SAMPLE.TXT" included on the distribution disk is to be checked. Any other file could be checked if desired. The 'C' command or "CONFIG.GMK" should not be used when first running Grammatik.

1. Start Grammatik by entering GMK after the CP/M "X>" prompt. Do NOT use the second command format (e.g., "GMK FILE").
2. After the command menu is displayed, enter "D" to read in the dictionary

"PHRASES.GMK".

3. Enter "I=SAMPLE.TXT" (or other filename) to set the name of the file being checked.
4. Enter "/" to start the checking process.

After Grammatik has been used a few times, other commands can be entered to exercise the different options available. All commands are terminated with the "RETURN" key.

4.1.2 -- Interactive command menu

The following command menu is displayed when Grammatik first starts up, and indicates the options that are available:

```

+-----+
:
: Aspen Software Grammatik (tm) CP/M V1.x
: (c)(p) 1981 Aspen Software Company
: Command,<>=optional / Meaning / Current setting
:
: C<=file> Read configuration file CONFIG.GMK
: D<=file> Read dictionary PHRASES.GMK
: E<=list> Errors, or NOE to exclude ABC ... XYZ
: F=char Set format command char
: I=file Set input file to check
: L<=file> List errors on file
: O<=file> Output with errors marked
: P or NOP Print errors on printer NOP
: S or NOS Show suggestions S
: T or NOT Show errors on TTY screen T
: W or NOW Wait when error displayed W
: // Done - begin checking
: (Use Control-C to abort)
:
: ***** Enter any command: _
:
+-----+

```

The left column of the menu contains the commands that may be entered. The middle column explains the command, and the right column shows the current setting or default for the command. After a command is entered, its effects will usually be reflected in the right column. Some of the commands (e.g., C<=file>) are shown with "<" and ">" indicating optional parts. If the part enclosed in "<>" is left out, then a default value shown in the right column will be used automatically. Note that the "<>" is not part of the command. The default for each command is explained with the command description below. The commands can be entered in any order, and some may be entered more than once as needed. Two of the commands, "C<=file>" and "D<=file>" cause immediate action, and may be used repeatedly before beginning to check the document. For example, the "D" command is used to read in a dictionary, and may be used repeatedly to read in as many dictionaries as needed. Most of the other commands set options that take effect only after checking of the file has commenced. The exception is "S", which affects both the output as the file is checked, and the method used to read in dictionaries with the "D" command. All commands must be

terminated with the RETURN key, and may be edited with the usual character delete key (DEL or BACKSPACE). Each command is described in detail below.

C<=file> - Read configuration file

This command may be used to read in a configuration file. A configuration file contains a list of commands to set various options. For example, the configuration file may contain a list of dictionaries to read in. If "C" is specified without giving a file name, then the default file name listed in the right hand column will be used. No configuration file is provided with the distribution package, and the "C" command should only be used after a configuration file has been defined as described in section 4.1.3. No configuration file will be used unless the "C" command is actually entered even though a file name is indicated in the right column.

D<=file> - Read dictionary

This command is used to read in phrase dictionaries. It is possible to read in as many dictionaries as are required for a given application. If the file name not specified, then the name shown in the right column will be used. It is possible to read in the same dictionary more than once, but double entries of each phrase will result. While any number of dictionaries may be read in, the total number of different entries from all dictionaries allowed is limited. These limits are specified in Appendix B. The "NOS" command can also be used to reduce memory requirements, and if entered before any dictionaries are read will cause the suggestion part of each phrase to be ignored.

The following example shows commands that could be entered to the command menu to read in the default dictionary "PHRASES.GMK" and another dictionary called "MYDICT.GMK". The parts enclosed in parentheses are comments, and would not be included on the command line. Note that the "D" command may be used repeatedly as needed to read in several dictionaries. A drive name may also be included if desired.

```
D          (read default PHRASES.GMK)
D=MYDICT.GMK (read in user dictionary)
```

No dictionaries will be read unless the "D" command is specified, even if the "Current setting" column indicates a file name. After a dictionary has been read, the default value will be cleared from the "current value" field. This prevents accidentally reading the same dictionary twice.

E<=list> - Errors, or NOE to exclude

This command is used to set which errors will or will not be displayed. By default, all errors will be shown. Some writers will find that some of the errors detected are more of a nuisance than a help. For example, this may be true for error type "M" - commonly misused words. The list supplied to the "E" command contains the errors to be displayed. For example, to display only errors of type "D" and "S", the command "E=DS" could be used. If just "E" is entered, then all errors from A to Z will be enabled. If "E=" with no list is entered, then all errors will be turned off. If the "E" is prefixed with "NO", then the errors in the list will not be listed. For example, the command "NOE=M" will turn off errors of type "M", while continuing to display all others. After the "E" or "NOE" command has been used, the "Current setting" column will be updated to reflect the errors that

will be displayed. Unlike the "C" and "D" commands, the "E" command (as well as all the rest) does not cause any immediate action. The options it sets are used after checking of the file begins.

F=char - Set format command char

Some word processors format text by embedding formatting commands within the text. A typical scheme used is to begin formatting commands on new lines that start with some special character. For example, Aspen Software's formatter Soft-Text includes formatting commands on lines that begin with a period (.). It is usually undesirable to include such formatting commands in the analysis of the document. The "F=char" command allows such lines to be ignored. WordStar and Soft-Text users should enter the command "F=." for example. This will cause all new lines that begin with the character "." to be totally ignored for analysis by Grammatik. If no format character is specified (the default), then all input text will be checked.

I=file - Set input file to check

The "I=file" command is used to specify the name of the file to be checked. The file name must be a legal CP/M file name. If the file name is specified when Grammatik is first invoked - "GMK FILENAME.EXT" - then the file name will be automatically be filled in and displayed in the right column. The file name that appears in the right column on the "I=" line is the file that will be checked, and is also used to build the default file names for the "L" and "O" commands. If no drive is specified, then the currently logged drive will be used. Files on any drive may be checked by giving a drive code with the filename. CP/M does not allow disks mounted in any drive to be changed once Grammatik is running, however. Using an extension when naming a file is optional, but usually useful to identify the type of information contained in a file.

L<=file> - List errors on file

This command is used to direct the error messages to a file. Only the error messages and the final summary are written to the file. No text from the source file is copied to the list file. Specifying "L" with no file name will cause the file to use the same name part as the current default input file shown after the "I" entry, but with an extension of ".LST". If "L=" (no name) is specified, then use of the list file will be cancelled. If no drive is specified with the file name, then the output will be written to the currently logged drive. The "L" and "O" commands are mutually exclusive; you cannot use both at the same time.

O<=file> - Output file with errors marked

The "O=" command is used to cause the input file to be copied to an output file with all errors marked. This is one of Grammatik's most useful options. The output file will be an exact copy of the original, except that all errors will be marked with three characters "?#x", where "x" is the error type. This marked file can then be edited with the word processor. Using a search command to find all instances of "?#" will show where all the errors are. The "?#x" can be deleted, and the text revised as needed. This process is especially easy if a list of errors was also printed on the printer. If both a marked output file is specified, and "W" wait switch is turned on, the user will be given the option of not marking the file when the wait message

appears. If the user just presses RETURN, the error will be marked in the output file. If the user presses the N key before the RETURN, the error will not be marked.

Soft-Screen users should note that the wild card character makes searching for the "?#x" especially easy. Enter a find pattern of "?#%", then use ";" to find each occurrence, and the "/" command to delete it.

Specifying "O" with no file name will cause the file to use the same name part as the current default input file shown after the "I" entry, but with an extension of ".ERR". After the ".ERR" file has been corrected, it should be renamed to the original name with the CP/M "REN" command. If "O=" (no name) is specified, then use of the marked output file will be cancelled. The "L" and "O" commands are mutually exclusive; you cannot use both at the same time. Warning: there must be enough free space on the disk to create a new file that will hold the marked copy of the original file.

P or NOP - Print errors on printer

This command controls error listing to the printer. If "P" is specified, then all error messages will be sent to the printer. You must have a printer connected and properly set up for this command to work! "NOP" (the default) will turn off printer output. This command is especially useful when used with the "O=" command, and will provide a list of errors and suggestions for use when correcting the document.

S or NOS - Show suggestions

This controls whether suggestions are included with the error messages. Most of the time, it is not desirable to turn off suggestions. If used before any dictionaries are read, the "NOS" command will cause the suggestion entry for each phrase in the dictionary to be ignored. This may be useful for systems with limited memory.

T or NOT - Show errors on TTY screen

This command can be used to turn off echoing of the source document on the screen. Note that the keyboard and screen are often called a terminal or TTY, and thus the use of "T" for this switch. If errors are being printed on the printer and the file copied to the marked output file, it may be desirable to turn off screen echoing. "T" turns on echoing, "NOT" turns it off.

W or NOW - Wait when error displayed

Normally, Grammatik will pause after each error is displayed on the screen. Checking will continue after pressing the RETURN key. If the "O=" option is turned on, the user will also be given the opportunity to not mark the output file (see the "O=" command description). There will also be the option of using "Q" to quit checking and return to the main command menu. "NOW" (no w) will turn the wait feature off, allowing Grammatik to check the file without any pauses.

It is also possible to temporarily stop the display at any time. Pressing the Control-S key will cause the display scrolling to stop until it is pressed again. This feature is often useful for halting the display from scrolling to read the rest of the context when continuing after an error message.

// - Done, begin checking

After all the desired commands have been entered, the "/" command begins checking. Checking will not begin until "/" has been entered, and no options may be changed once checking has begun. No checking for a correct input file name is done until the "/" command is used. If no input file name has been specified, or if an incorrect file name is supplied, Grammatik will request that the user enter a correct name.

4.1.3 -- Automatic Configuration

Entering each command every time when using Grammatik can become tedious. Therefore, Grammatik provides a mechanism for automatically setting options and reading dictionaries. When Grammatik is invoked and a file name is specified (e.g., "GMK FILENAME"), Grammatik will automatically search for a file called "CONFIG.GMK". That file should contain commands to set the options in the same form as would be used when setting them interactively.

The file "CONFIG.GMK" can be created with ED or the user's own word processor. It should be saved as a standard ascii file. (WordStar users should use the non-document editing mode to create a configuration file.)

Suppose as the normal checking procedure it is desired to read in the two dictionaries PHRASES.GMK and SEXIST.GMK, set the format character to a period, list error messages on the printer, and copy the source file to the marked output file. Using either the user's word processor or the CP/M editor ED, a standard ascii text file called CONFIG.GMK could be created containing the following five lines:

```
D
D=SEXIST.GMK
F=.
P
O
```

Note that each command must be on a separate line terminated with a standard CP/M end of line mark (carriage return - line feed), usually entered with the RETURN key.

Suppose Grammatik is then invoked with the command:

```
GMK MYFILE.TXT
```

First, Grammatik automatically sets the input file to "MYFILE.TXT". It then checks for CONFIG.GMK. It finds it, and reads in commands from the file, just as if they had been entered from the keyboard. The first line in the example causes the default dictionary PHRASES.GMK to be read in. The second line causes SEXIST.GMK to be read. The third line sets the format character to ".", useful for Soft-Text and WordStar users. The next line turns on the printer listing. The last line causes the input to be copied to an output file with errors marked. In this example the output file is called MYFILE.ERR. Once all these commands have been processed, Grammatik then begins to check the file immediately with no additional command input. No "/" is necessary, since the end of the file in CONFIG.GMK indicates the same thing.

If the configuration command file contains any errors, Grammatik will abort reading it, and return to the interactive command menu. This can be handy if an illegal command (such as X) is purposely included as the last command of the configuration file. This automatically sets up the most commonly used options, yet still allow commands to be entered from the interactive menu. Using the "C" command within a configuration file is illegal, and will cause an error abort to the menu.

Another example:

```
D=A:PHRASES.GMK
F=\
E=M
X
```

The first line reads the dictionary from drive A:. The second line sets the format character to "\", useful for Magic Wand users. The third line turns off only "M" errors. The last line is an illegal command, and causes Grammatik to return to the interactive command menu.

If Grammatik is invoked using the "/B" switch (e.g, "GMK MYFILE.TXT/B"), it will try to read a bad word file called "MYFILE.BWD" before reading the configuration file. This option is intended so that Proofreader may automatically pass a bad word list to Grammatik.

It is probably easiest to use Grammatik from drive A:, and keep the documents to check on B: or some other drive. The master dictionaries "PHRASES.GMK" and "SEXIST.GMK" can also be kept on A:. Start Grammatik by first logging on to drive A: with the "A:" command. The automatic configuration file should also be kept on A:. Then Grammatik can be invoked with "GMK B:FILE.TXT".

4.2 PROFILE

PROFILE will produce a profile of word usage in a document. Start PROFILE by entering "PROFILE" after the CP/M "X>" prompt. The program will then request the name of a file to profile and give an opportunity to direct the profile output to a file or to the printer. The operation is simple. The profile produced begins with a line indicating use count, followed by all the words used that many times. The least used words (e.g., 1 time) appear first, and the most used words last. Words that appear only once are most likely to be misspelled.

4.3 SORTDICT

The SORTDICT utility is used to sort phrase dictionaries. While Grammatik does not require that dictionaries be sorted, they are easier to read and maintain if they are. Start SORTDICT by entering "SORTDICT" after the CP/M "X>" prompt. The program will then request the name of a file to sort. The size of a file that may be sorted by SORTDICT is limited by RAM size. This should not be a practical limitation since Grammatik also requires that dictionaries fit into RAM. SORTDICT knows about the format of a dictionary file, and will sort in alphabetical order on the phrases only. Either the "*" character used to mark the end of a phrase in the dictionary, or an end of line (RETURN) will delimit the part of a line used in the sort. SORTDICT may be used to sort almost any file with data items contained on separate lines.

SORTDICT will read in a file, sort it, and then write it out back to the same file. A file should never be sorted unless a backup copy of it has been made first!

Appendix A -- Error messages

The errors detected by Grammatik have been divided into several categories or types. Many of the error types, such as "D" or "U" almost always indicate an error. Others, such as "W" or "O", indicate a potential style problem that is subject to personal taste. Any error types not used by Grammatik can be used for user defined error types when building additional dictionaries.

A - Archaic usage

This error type is used for words that are archaic, and have fallen out of common usage.

B - Unbalanced {}[] or "

This message will appear at the end of the analysis phase, and indicates that either there was not a matching right paren for each left paren, or that there were an odd number of double quotation marks.

C - Capitalization error

Grammatik checks for consistent capitalization. Words that begin sentences and the single letter "I" must be capitalized. Other words must be either all lower case or all upper case, or have a capital letter only as the first letter of the word.

D - Doubled word or punc

Indicates that a doubled word ("the the") or punctuation mark has been detected. Punctuation checked for doubling includes: "?", ".", "!", ",", and ";". All other characters are not checked for doubling since it is often intentional in those cases.

E - Error - unspecified

This category is reserved for user dictionaries. No entries in the dictionaries provided with Grammatik use "E".

G - Gender specific term

The words provided in SEXIST.GMK use error message 'G'. Writers trying to remove terms that are gender specific, and therefore potentially sexist will find this error category useful. Sometimes it may seem difficult to remove such terms, but it is possible. The reader should note that this manual has been written without the use of gender specific terms.

I - Informal usage

Some words, such as "ain't", do not belong in formal writing. These have included as type "I" errors.

J - Jargon or technical

No entries provided in the default dictionary use error type "J". However, the user is free to use "J" to mark errors when building dictionaries of esoteric jargon or technical terms. This category can be especially useful when writing documents intended for a general readership.

K - Awkward usage

Some phrases, such as "and/or" are awkward, and best revised.

M - Commonly misused word

Some writers have trouble when using similar words, such as eminent, imminent, and immanent. The dictionary contains examples of several

commonly misused words of this type. The suggestion provided with these words gives the alternate words, and a definition of the word detected. Writers who are comfortable with their usage of such words may want to use the "E" command to disable "M" error detection.

O - Overworked or trite

Some phrases, while not being wordy or incorrect, are overworked. The overworked phrase should be removed and the sentence revised.

P - Punctuation error

Grammatik can detect several punctuation errors, including missing spaces after a punctuation mark, and placement of end of sentence punctuation outside quotation marks.

R - Redundant phrase

Many phrases, such as "seldom ever" or "join together" are redundant, and should usually be revised.

S - Spelling error

Grammatik can detect some spelling errors, such as "can not" that will not be found by a spelling checker.

T - Trademark

Certain words, such as "Xerox", are trademarks.

U - Improper usage

Grammatik will detect some phrases that are used improperly, such as "must of" instead of the correct "must have".

V - Vague adverb

Often a more precise word can be used.

W - Wordy phrase

The largest error category contained in the default dictionary is "W". Phrases such as "all of" or "a number of" can usually be replaced with more concise terms such as "all" or "several".

Appendix B -- Dictionary structure

Phrase dictionaries are standard ASCII text files. Dictionaries may be edited or created using either the user's word processor or the CP/M editor ED. (WordStar users use non-document mode.) They may be listed using the CP/M "TYPE filename" command.

The total number of different entries allowed from all dictionaries is 1000. The default dictionary "PHRASES.GMK" contains just over 500 entries, the "SEXIST.GMK" dictionary contains about 100 entries, leaving 400 additional entries for the user if both default dictionaries are used. The total number of characters allowed in the dictionary definitions is also limited by RAM size.

Each individual entry in a phrase dictionary must be placed on a separate line. Each line must be ended with an "end of line" character, which is usually entered with the "RETURN" key. Dictionaries do not need to be in alphabetical order. Duplicate entries are not fatal, but will cause repeated error messages to be displayed when that error is detected.

Each entry in the dictionary has the following format:

```

without further delay#w>now, immediately
      :           :           :
- phrase -----+           :           :
- end of phrase mark -----+!           :           :
- error type -----+!           :           :
- beginning of suggestion ----+           :           :
- suggestion -----+           :           :

```

The first entry is the phrase or word to be detected. The phrase may contain any character other than "#". The end of the phrase is marked either by the end of the line or the character "#". If a "#" is used, then an error type or user category number will be expected after the "#". If the phrase is ended by the end of line, then the phrase will automatically be assigned to error type "E" and not use a suggestion. It does not matter if upper or lower case is used for dictionary entries since Grammatik knows about case, and will ignore the difference between upper and lower case.

The error type must be a single letter from A to Z. The user may use any of the letters, even if no error message has been assigned to that letter. If the entry is to be a phrase or word in a user category, the error type should be a single number from 3 to 9.

Following the error type comes the suggestion. If no ">" follows the error type letter, the phrase will not have a suggestion listed when it is detected. The suggestion may consist of any words or characters desired. The entire line must be terminated with an end of line (RETURN).

User categories 3-9 should not use a suggestion. For example, user category entries would appear in the following format:

```

that#3
which#4
. And#5

```

Categories 3 and 4 count occurrences of "that" and "which", while category 5 counts sentences that begin with "And". (When matching patterns, any number of consecutive blanks, tabs, or end of lines are counted only as a single blank.)

Appendix C -- Special words

Grammatik recognizes three special categories of words: abbreviations, to be's, and common prepositions. The recognition of abbreviations is fixed, and built into Grammatik. To be's and prepositions are contained in the default dictionary, and may be changed by the user.

C.1 - Abbreviations

The following words are recognized as abbreviations by Grammatik. If it is desired to check for any of these abbreviations in the dictionary, the period at the end should be included. All single letters followed by a period are also treated as abbreviations by Grammatik.

al. av. ca. co. ct. cu. dr. et. ft. hr. in. jr.
kg. km. lb. lt. mi. ml. mm. mr. ms. mt. no. oz.
ph. rd. rt. sc. sq. sr. st. vs. wk. wt. yd. yr.
adm. adv. apr. aug. ave. col. dec. drs. etc. feb.
fig. gal. gen. gov. hrs. inc. jan. jul. jun. kgs.
kms. lbs. maj. mar. mgr. min. mrs. nov. oct. ozs.
pvt. rep. rev. sec. sen. sep. sgt. viz. yds. yrs.
atty. blvd. capt. corp. dept. gals. govt. mins.
pres.

C.2 - To be's

am are be been being is was were

C.3 - Common prepositions

as at by for from in like of on to with

Appendix D -- Bibliography

Following are a few references that may be of additional value when revising documents. This list is by no means comprehensive.

A Manual of Style. Chicago: The University of Chicago Press, 1969.

Brusaw, Charles T., Gerald J. Alred, and Walter E. Oliui. Handbook of Technical Writing. New York: St. Martin's Press, 1976.

Hodges, John C., and Mary E. Whitten. Harbrace College Handbook. New York: Harcourt Brace Jovanovich, 1977.

Lanham, Richard A. Revising Prose. New York: Charles Scribner's Sons, 1979.

Montgomery, Michael, and John Stratton. The Writer's Hotline Handbook. New York: New American Library, 1981.

Strunk, William, and E. B. White. The Elements of Style. New York: Macmillan Publishing, 1979.

INDEX

abbreviations - 7, 22
 aborting checking - 8
 automatic configuration - 11, 16

Backing Up - 4

c - read configuration file - 13
 capitalization - 3, 8
 checking another document - 10
 checking procedure - 7
 command menu - 12
 command options - 7, 8

d - read dictionary - 13
 dictionary - 7
 dictionary structure - 21
 distribution package - 4
 doubled words - 3

e - set errors - 13
 error messages - 8, 19
 exiting - 10

f - set format character - 14

halting scrolling - 8

i - 8
 i - set input file - 14
 introduction - 3
 invoking grammatik - 11

l - list errors on file - 14

Magic Wand - 17

marking errors - 8, 13, 14

new document - 10
 numbers - 7

o - marked output file - 14
 organizing files - 5
 overview - 6

p - print errors on printer - 15
 parentheses - 3, 7
 prepositions - 9, 22
 profile - 3, 6, 18
 Proofreader - 11, 17
 punctuation - 7

quotation marks - 3, 7

readability - 9

show suggestions - 15

Soft-Screen - 15

Soft-Text - 14, 16

sortdict - 18

start checking - 16

starting grammatik - 11

starting up - 11

suggestions - 8, 15

summary - 9

t - shows errors on terminal - 15

to be's - 9, 22

typographical errors - 3

undisplayable characters - 8

user categories - 9, 21

w - wait when error - 15

word - 7

WordStar - 11, 14, 16, 21

writing style - 3

ASPEN SOFTWARE LIMITED WARRANTY AND LICENSE

Aspen Software Company programs are warranted to be free from copying defects at the time of shipment. If the copy of a program will not load or run, it may be returned for free replacement for up to two weeks after receipt. Aspen Software will also reply to telephone inquiries regarding any problems with a program. The Aspen Software Company is not responsible in any way after the first two weeks. Aspen Software will notify registered licensees of any known bugs or problems in the program for a period of at least six months.

All Aspen Software programs and documentation are distributed for the sole use of the original licensee on a single computer. Copies of the machine readable programs may be made only for legitimate backup purposes of the licensee, and may not be sold, given, loaned, or in any other way conveyed to any other person. All Aspen Software programs and documentation are copyrighted materials, and all rights of reproduction are reserved to Aspen Software Company.

Registered licensees of Aspen Software Company programs will be entitled to low cost upgrade service when major new versions of the programs are released. This upgrade service will require the return of the original distribution disk. There will be a \$10.00 to \$50.00 upgrade fee, plus any difference in the price of the new version and the old version. Only those licensees who bought the software directly from Aspen Software originally or those who return the registration card provided with programs purchased through dealers will be entitled to the Aspen Software Company warranty and upgrade service.

There are no other warranties, expressed or implied, including, but not limited to, any implied warranties of merchantability or fitness for a particular application. All Aspen Software programs and documentation are sold on an "as is" basis. All programs have been thoroughly tested, but there may be bugs for which Aspen Software cannot be responsible. Aspen Software Company assumes no liability or responsibility to the customer or any other person for any application or use of any software or documentation, sold by it or any of its dealers.

TRSDOS, TRS-80, and SCRIPSIT are registered trademarks of Tandy Corporation. CP/M is a registered trademark of Digital Research. WordStar is a registered trademark of MicroPro. Magic Wand is a registered trademark of Peachtree. PROOFREADER, Proof-Edit, Soft-Screen, Soft-Text, Grammatik and Aspen Software Company are registered trademarks of Aspen Software Company, a New Mexico Corporation.

(c) Copyright 1981,1982 Aspen Software Company.
All right reserved.