LEXICO Guide No. 1
OVERVIEW

Richard L.Venezky
Nathan Relles
Lynne A. Price

Computer Sciences Technical Report #282

December, 1976

LEXICO Guide No. 1

OVERVIEW

Richard L. Venezky
Nathan Relles
Lynne A. Price

TABLE OF CONTENTS

1. Introduction

1.1 System Capabilities

LEXICO is a text processing system with the following capabilities:

(1) forming and maintaining a collection of texts, including entering new texts into a collection, deleting and editing texts already in a collection, and specifying collection parameters to reduce repetition of control statements;

(2) concording individual texts (listing each word of the text with the contexts in which it occurs);

(3) classifying words in a text by headword (also called base form or lemma);

(4) generating slips for a dictionary file or a base concordance (listing each headword with the contexts in which its text forms occur).

The system was designed primarily for use in lexicographic work, but has other applications. At the University of Wisconsin, where the system was designed, it runs on a UNIVAC 1110 with any teletype-compatible terminal. Since coding was done in FORTRAN, however, conversion to other machines with an identical word size (six 6-bit characters per machine location) is possible.

1.2 Components of a Collection

All text processing is centered around the user's collection. A collection is composed of texts which the user wants to process in similar ways; for example, all the texts whose words will be classified under one set of headwords. In addition to the texts themselves, a collection contains a directory identifying and describing each text, and may contain any of the following:

(1) conventions to be used when entering or concording texts;

(2) spelling conversion rules to be used for headword classification;

(3) a text form--base form conversion list to be used for headword classification;

(4) concordances of texts;

(5) for each concorded text, a list of words occurring in the text.

Features which may be common to all texts in a collection are called collection defaults. Once specified for a collection, these features need not be respecified for each text.

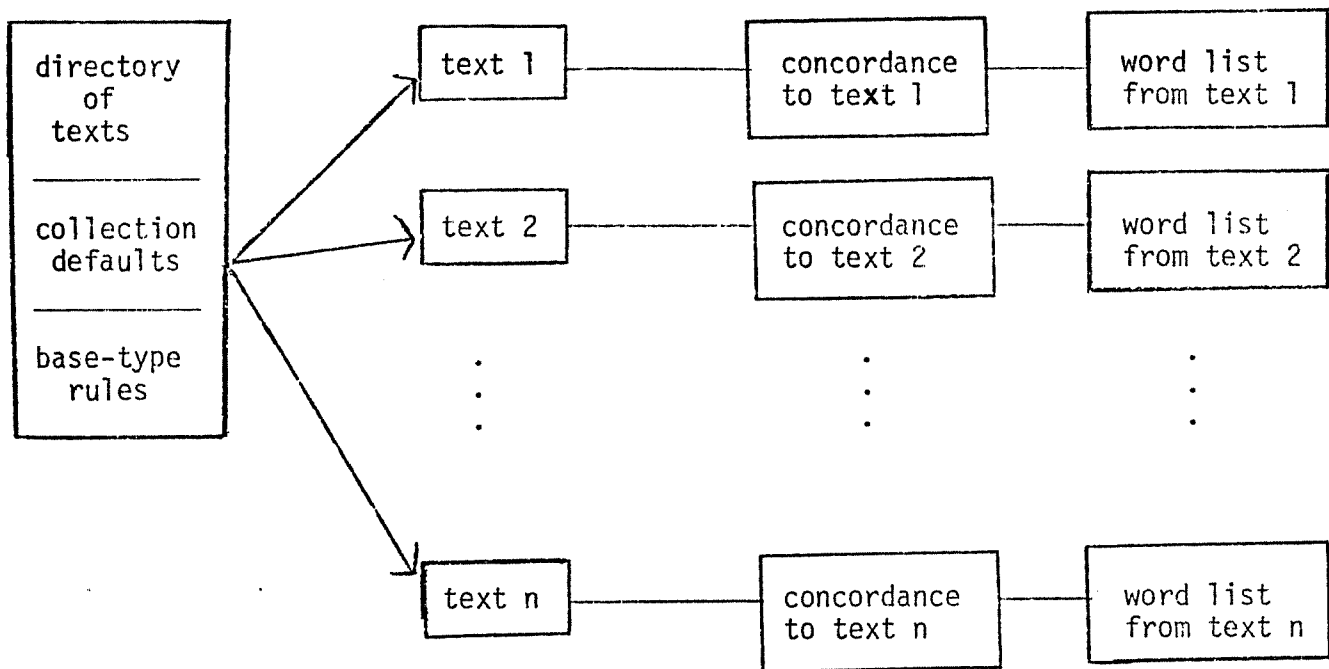Figure 1 illustrates the components of a collection which are accessible by the user.

Figure 1

Components of a Collection which are Accessible
to the User

1.3 Major Processes

The typical processes which can occur for a collection of texts forming the corpus for a dictionary include:

(1) creation of a collection and, optionally, definition of collection default values;

(2) entry of one or more texts into the collection, optionally producing a listing of each text;

(3) correction of errors in texts (editing);

(4) concording texts;

(5) formation of a headword classification (lemmatization) process for the collection by definition of spelling conversions and base form--text form associations;

(6) headword classification of the words that occur in texts, producing, for each text, a listing of the headword (base form) associated with each text form;

(7) alphabetization of a concordance according to the spellings of the base forms;

(8) generation of slips;

(9) deletion of texts from the collection; and

(10) deletion of the collection from the system (after all processing is complete).

In addition, users may at any time examine a summary (the collection directory) of the processes performed on each text. Users may also obtain

listings of any text, of the spelling conversion rules and base form--text

form associations, or of the base forms associated with each word in a text.

The latter listing may be generated showing the contexts in which forms of

each base occur. Also, collection default values may be inspected or updated;

the headword classification process may be revised; and the text form--base

form assocations of any text may be corrected.

1.4 The LEXICO Guides

Documentation for LEXICO is arranged in a sequence of guides, each

describing different components of the system. These are:

Guide No.	Title
1	Overview
2	Communicating with LEXICO
3	Creating and Maintaining a Collection
4	Preparing and Entering a Text
5	Editing
6	Concording
7	Headword Classification

Sections 2-6 of this overview briefly introduce features which are

described in detail in the other guides. Section 7 contains a glossary of

the terminology used throughout the LEXICO system.

2. Communicating with LEXICO

2.1 Introduction

This section describes how a user interacts with LEXICO and system capabilities that facilitate this communication. A detailed description of these conventions is presented in Guide 2.

The user enters commands to LEXICO from a remote terminal. Many tasks are performed on-line; the results are displayed immediately at the same terminal. However, some tasks require so much time, generate so much output, or cost so much, that they are perfomed off-line. In this case, the user enters commands at the terminal and the jobs are recognized at once, but the tasks are actually performed later and results are printed at the computing center.

A dialogue between a user and the LEXICO system consists of:

(1) commands by the user to the system and responses to those commands;

(2) questions asked of the system by the user and the system's responses;

(3) questions asked by the system of the user and the user's responses; and

(4) error messages displayed by the system when a command is not understood.

2.2 The LEXICO Command Language

Instructions to LEXICO are entered in the form of commands (also called statements). The construction of each command is described with the task to which it pertains, but some examples are given here:

```
CREATE; PHYSICAL LENGTH 80;      COLLATE    STANDARD ;
ADD STOPWORDS AUS AUSSER BEI
MIT NACH  SEIT
VON ZU ;
```

Note that each command is terminated with a semicolon. A command may extend over several lines, or several commands may be entered on one line. Also, any number of blanks may occur between command language words.

Many commands have abbreviated forms; for example, SHOW DIRECTORY and SH DIR are equivalent.

After some commands are entered, LEXICO asks a question of the user; for example, the command

        EDIT ;

is followed by the prompt: WHICH TEXT?.

Most tasks are specified in the system as functional blocks. A block begins with a block header, a command which identifies the task (e.g., CREATE, ADD, EDIT). Subsequent commands specify how the designated task is to be performed (e.g., delimiter specifications, printing options, collating sequence). Following these commands and declarations the statement

        END ;

is entered. If the block has been entered for testing purposes, or if errors exist which cannot be corrected immediately, the command

        IGNORE ;

may be used instead of the END statement. This nullifies any specifications made in the block.

Some tasks may be completely specified in one command (e.g., SHOW DIRECTORY). These single-statement blocks do not require END statements.

The following is an example of a block entered to add a text to a collection and to generate a concordance of that text. The indentation of commands is not required, but is done here to distinguish among the block header, END statement, and parameter declarations.

```
ADDCONCORD 'BEOWULF V.L' ;
  CITATION DELIMITERS # ! . ;
  INPUT ON TAPE 1234 ; ADD STOPWORDS
  'AND'    'BUT' 'OR' 'FOR' 'NOR'
  'EITHER' 'NEITHER' ; SEQUENCE ON '%' ;
  TEXT OUTPUT ON PRINTER ;
END ;
```

2.3 On-line Aids

During interaction, the user may enter special commands to LEXICO for assistance in understanding error messages, questions, or required information. These capabilities are especially suited for, but not limited to, the novice user.

In general, each of these commands is a three-letter abbreviation preceded by an asterisk. On some terminals, a single key may be pressed to get the same effect. (On the Hazeltine 2000 terminal, these keys are located on the right-hand side of the keyboard, and must be pressed while the CTRL key is held down).

The aid commands, which are not terminated by semicolons, are:

Explain Error (*ERR). If an error is made, the system displays a brief description of that error. In response to *ERR, a more detailed explanation is displayed. *ERR may be entered several times for progressively more detailed explanations.

Explain Question (*EQU or *EXQ). Occasionally, the system will solicit information in the form of a brief question. By successively entering *EQU (or *EXQ), progressively more detailed explanations of the soliticed information are displayed.

Example (*EXA). Whenever the system solicits information or when an error has been made on previous input, examples of correct input are obtainable with this command.

Help (*HLP). By entering *HLP, a user can obtain help of a general nature, viz., a description of other user aids and, in the case of errors, corrective measures.

Menu (*MNU). Only certain commands are allowed within a given block (e.g., editing commands may be entered only in an EDIT block). The special code *MNU causes the system to display all commands allowed at any given time.

Cost (*CST). Entering *CST causes LEXICO to display the cost so far accumulated by the current interactive session, along with the dollar limit for the run. If *CST is entered more than once during a single session, the cost accumulated between its occurrences is also displayed.

Comment (****). Any line beginning with four asterisks is ignored by LEXICO. This command is useful for documenting or annotating an interactive session with the system.

2.4 A Sample Dialogue

The following dialogue between a user and the LEXICO system illustrates some of the features described above. User input is in lower case and is proceded by the symbol >. System messages are in the upper case.

Only the first input requires special explanation. The procedure for signing on and initiating interaction with the central computer site differs from one user to another, depending on the terminal characteristics, user's project (account) number, password, etc. Once in contact with the computing system, @LE*ICO. is entered to initiate interaction with LEXICO.

> @le*ico.

LEXICO VERSION 1.3 05/29/76 13:00:06
COLLECTION NAME?

> *exq

PLEASE ENTER THE NAME OF THE COLLECTION YOU WISH TO ACCESS OR ARE ABOUT
TO CREATE FOR THE FIRST TIME.

> *exq

A COLLECTION NAME CONSISTS OF ELEVEN OR LESS CHARACTERS, ALL OF WHICH
ARE ALPHANUMERIC (A-Z, 0-9).

> poe

COLLECTION ACCESSIBLE; PROCEED.
TASK COMMAND ;

> *exq

YOU MAY NOW ENTER A BLOCK HEADER, OR TASK SPECIFICATION, IN THE COMMAND
LANGUAGE.

> edit nevermore ;

NEVERMORE ?
NO SUCH TEXT .

> *exq
NO QUESTION TO EXPLAIN.

> *err

NO TEXT WITH THE ABOVE SPECIFIED NAME EXISTS IN THIS COLLECTION.

> *err

THE DIRECTORY OF TEXT NAMES IS OBTAINABLE BY ENTERING:
        SHOW DIRECTORY ;

> show directory ;

TEXT CODE	TEXT NAME	CIT	CON	E	R	L	S
1	CASK	143					
3	RAVEN	18					
7	USHER	168					

> **** who changed the name of this text???

> edit raven ;
TEXT RAVEN (TEXT CODE: 3) AVAILABLE:

3. Preparing and Entering a Text

3.1 Introduction

This section contains a general description of the components of a text. Detailed instructions on the preparation of a text for entry into a collection are contained in Guide 4.

3.2 Citations

All texts for a collection are divided into citations. These are user defined sytactic/semantic units. They may be phrases, physical lines, sentences, or sentence sequences. When a text is entered, the boundary of each citation is unambiguously marked by a special terminating character called a citation delimiter. Citation delimiters may be conventional punctuation (.?!), or any other user-defined characters.

3.3 Notes

Each citation may have associated with it material which is not part of the citation proper and which would not normally be concorded. This might, for example, be a translation, source description, or variant reading. Such material, called a note, is enclosed within note delimiters.

3.4 Identifiers

Finally, every citation has an identifier (ID) associated with it, such as page and line numbers or psalm and verse numbers. Identifiers may have up to three levels and each level may be alphabetic or numeric.

Levels are separated by a special character selected by the user called the ID level delimiter. The entire identifier is enclosed within ID delimiters. When entering a text, it is possible to omit the identifier before some citations-- LEXICO will construct one based on the ID of the previous citation. However, an ID must appear before the first citation of each text.

Some examples of IDs are:

```
[1/3]
(11.D.6)
/3-2/
+09/03/75+
=10:23:2=
```

3.5 Text Names and Delimiters

All texts must begin with a <u>text name</u>, enclosed in single quotes. A

text name is a string of up to 12 characters. A text name may not begin with

a digit (0-9) and cannot contain a quote mark ('). All texts must end with a

special symbol, defined by the user, called the <u>text delimiter</u>.

3.6 Example of a Coded Text

In the following example, IDs are enclosed in square brackets and notes

between equal signs (=). Ordinary punctuation is used to mark citations, while

the entire text is delimited by a number sign (#). Note that the translations

of the Latin comments are delimited as notes. The asterisks are substitutes

for the symbol thorn.

```
'Chaucer ABC'

[1]   Incipit carmen secundum ordinem litterarum
      alphabeti = Here begins a poem following the order
      of the alphabet=.

[2]   Almighty and al merciable queene,
      To whom *at al *is world fleeth for sucour,
      To have relees of sinne, of sorwe, and teene,
                      .
                      .
                      .
[184] To penitentes *at ben to merciable, amen.
[185] Explicit carmen. = Here ends the poem. =#
```

3.7 Adding Texts to a Collection

Once a text has been coded and transformed into machine readable form, it must be stored on a disk file (see Guide 4). The user then requests that the text be added to a particular collection, and at the same time makes all text declarations that are not already implied by the existing default values. Control statements exist for entering one or more texts at a time, and for entering and concording a text (or texts) in a single job.

4. Editing a Text

An edit block is used to edit a text in a collection. Facilities are provided for

(1) inspecting the current contents of a text ;

(2) inserting, deleting, or replacing citations ;

(3) combining or splitting citations ;

(4) altering portions of a citation, or a sequence of citations; and

(5) displaying or changing citation identifiers (IDs).

During the editing session, a pointer indicates the 'current citation'. Requested changes are made relative to this citation. For example, citations may be inserted after it or changes made to it. In addition to the commands for modifying the text itself, several EDIT commands may be used to move the pointer from one citation to another.

Complete specification of editing commands is given in Guide 5.

5. Concording a Text

5.1 General Description

An alphabetized listing of each unique word, or type, that appears in a text, with the citations in which each word occurs, is called a concordance. There are two classes of types entered into a concordance-- keywords, listed with their citations, and stopwords, listed only with the number of times (frequency) they appear. The user may specify various rules for determining which words in a text are classified as keywords and which as stopwords. Concording options are described thoroughly in Guide 6.

5.2 Keywords and Stopwords

As a concordance is generated, each word in the text is tested to determine whether it should be classed as a keyword or as a stopword. A user may provide the following as a basis for this distinction:

(1) stopword list--a list of words to be classed as stopwords, either completely specified by the user, or based on a list of standard stopwords provided by LEXICO ;

(2) stop characters--characters which, when they appear at the beginning of a word, designate that word as a stopword ;

(3) stoppable lengths--designation of words of a particular length as stopwords; and

(4) reversability--specification that all words that would otherwise be classed as stopwords should be treated as keywords and vice versa.

5.3 Other Concording Options

The user may also specify

(1) removable characters--characters which are stripped off the beginning or end of a word before that word is entered in the concordance ;

(2) a <u>collating</u> <u>sequence</u> or alphabetical order to be used in arranging keywords and stopwords ;

(3) a physical device other than the printer to which concordance output should be directed; and

(4) a limit on the number of citations to be concorded.

5.4 Initiating a Concordance

A concordance of a text that has already been entered into a collection, and perhaps been edited, is scheduled with a CONCORD block. A text can be concorded at the same time it is entered into a collection with an ADDCONCORD block. There is no limit to the number of times a text may be concorded.

6. Headword Classification

6.1 Introduction

This section briefly describes lemmatization and the generation of slips. Lemmatization or headword classification is the association of a base form with each word that occurs in a text. For example, 'be' may be considered the base form for 'is, am, were ...'. A slip lists a single citation from a concordance, along with one of its text forms, the corresponding base form, and source identification. One slip is generated for each concordance entry. The commands and blocks used to perform these processes are explained in Guide 7. The tasks are performed on a word list which is produced whenever a text is concorded. This list contains all the keywords and stopwords which appear in the text.

6.2 Respelling

The user may wish to standardize the spellings of words by the application of spelling rules. For example, in Old English texts, eth and th can be converted to thorn. If this is done, several entries in the word list may have the same respelled form. Since the user specifies the base form to be associated with a respelled from, respelling allows him to reduce the number of forms that must be matched. In general, a spelling rule has the form r : s where each of r and s is a string of up to five characters. This means that the string s is to be replaced by r wherever the former occurs in the word list. No change is made to the body of the text. Erroneous forms generated by the application of spelling rules can be corrected in a CLEANUP block.

6.3 Basetype Rules and LOOKUP

LEXICO permits the user to associate a base form with each entry in a word list in either of two ways. He may specify the bases on-line, or he may request that previously declared basetype rules be applied to the word list of a text. The advantage of the latter process, called LOOKUP, is that a single set of rules can be used for all texts in a collection. This set of rules may be applied to the word list of any text. The rules have the form **base:type** where base is the base form for type. If respelling is used, type is a respelled form; otherwise, it is an original form.

6.4 CLEANUP

One of the uses of the CLEANUP block is to declare interactively the base forms of entries in a word list. LEXICO will display each original form (with the respelled form, if respelling is used) and ask the user to enter the base form. It will then formulate the corresponding basetype rule and ask the user if he wishes to have it saved in the collection. Rules that are generated in this fashion can be applied later in LOOKUPs of word lists of other texts.

Other functions of the CLEANUP block include the correction of errors generated by RESPELL and LOOKUP; entry of bases for forms left unmatched during LOOKUP; removal of unwanted types, such as numerals, from the word list; and the association of citations with homograph bases.

6.5 Homographs

LEXICO allows the user to designate a spelling as a homograph and to identify the correct base for individual occurrences of such a word.

A special kind of basetype rule called a homograph rule permits as many as seven different bases to be associated with one type during LOOKUP. Homograph bases may also be declared in CLEANUP. In CLEANUP, the user associates each occurrence of the word with one of the bases.

6.6 Slips

After bases have been associated with each entry in a word list, LEXICO will alphabetize a concordance by the spellings of these headwords. This information can be obtained in any of three ways:

(1) slips--each 5½" X 6½" page contains a base form, text form, a single citation, and the text and collection names ;

(2) base concordance--a listing similar to that produced by a concordance with the entries alphabetized by the spelling of the base form and with each text form given with its respelled form (if any) and base form; and

(3) slips file--a user may wish to have the concordance printed in a format different from any of those just described. LEXICO will produce a data file which the user can then reformat with his own program.

7. Glossary

alphabetic ID

An ID in which every level begins with a letter and contains only letters and digits.

alphabetic ID level

An ID level consisting of letters and digits and starting with a letter.

alphanumeric character

Letter or digit.

autostop

Character which, when appearing at the beginning of a word, designates that word as a stopword.

backup

A copy of a collection kept in case a system error destroys the actual collection.

bpi

Bytes per inch (the maximum number of characters that can be stored on one inch of tape).

base

Canonical form of a word; also called entry form, lemma, or headword.

base concordance

One form of SLIPS output--a concordance in which the keywords and stopwords are alphabetized by base form and are listed along with their text form, respelled form (if any), and citations.

basetype

Pertaining to association of a base with a text (or respelled) form.

basetype rule

A rule for classifying a text form (or respelled form) with a base form.

batch

    Off-line

blocked records

    An efficient method of storing data on a computer tape in which several
    units of information are written at one time in order to reduce the
    number of spaces that must be left between groups of data on the tape.

block header

    The first command in a block, specifying the task to be performed, and,
    for text-specific blocks, the text on which the task should be performed.

bottom

    The last citation in a text (used in EDIT blocks).

CLEANUP

    A block used for making changes in the word list of a text and for assigning
    citations to homograph bases.

citation

    A user-defined unit of a text such as a phrase, physical line, or sentence
    (sometimes used to mean the portion of text defined above and the associated
    ID and notes, if any).

citation delimiter

    A one-or two-character symbol used to mark the end of a citation.

collating sequence

    'Alphabetical order'--the order in which characters should be arranged when
    sorting for a concordance or when generating slips.

collection

    A group of texts along with a directory of the texts and various collection
    defaults, rules for headword classification, and possibly, concordances
    and word lists of texts.

collection default

>   A parameter value which holds for all texts in a collection unless
>   explicitly overridden.

command

>   An instruction to the LEXICO system.

concordance

>   A listing, arranged in alphabetical order, of each word that occurs
>   in a text with all its contexts.

control statement

>   Statements required by the UNIVAC 1110 operating system for identifying
>   a run, establishing files, etc.

COPY,G format

>   A UNIVAC 1110 format for copying disk files onto tape. See MACC Computing
>   Handbook for technical details.

crash

>   A state of primordial chaos caused by an unexpected and completely
>   disastrous malfunction of the computer.

current citation

>   In an EDIT block, the citation to which the pointer is currently positioned.

default

>   A value which holds for a parameter when no other value is overtly defined.

delimiter

>   A one-or two-character symbol used to mark the boundary of a unit of text
>   (e.g., ID, citation).

digit

>   Any of the characters 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

directory

> A list of the texts in a collection showing the processes that have been performed on each.

disk file

> A file stored on a disk, as opposed to tape. A disk is a rotating mechanism that allows reasonably quick access to any segments of a file. LEXICO disk files are always on-line files.

display (v)

> As used by LEXICO, to show information at an on-line terminal.

dollar limit

> The maximum amount of money which a user is willing to pay for a job. The UNIVAC 1110 does its accounting while a job is running; therefore, if this limit is exceeded before the job is completed, processing terminates prematurely. Under such conditions usually no useful output is obtained.

endstrip

> Characters which occur in a text, but are removed from the ends of words when forming concordance entries.

entry form

> Base.

file

> Any accumulation of data or programs which is overtly delimited(and named) according to UNIVAC 1110 file conventions. LEXICO creates certain files for the user; others can be declared by the user himself.

file charge

> A daily charge by MACC for maintaining a disk file. This charge includes a fixed fee for each file, plus a variable charge based on the amount of disk space which the file requires.

frequency

> The number of tokens of a type.

frontstrip

Characters which occur in a text, but are removed from the beginning of words when forming concordance entries.

headword

Base.

high density

800 bpi (for 7 track tapes).

homograph

A text form that has two or more bases, such as English lead or wind.

homograph base

One of the base forms associated with a homograph.

homograph rule

A basetype rule used to associate several bases with one text (or respelled) form.

ID

Identifier of a citation, such as page and line numbers or psalm and verse numbers.

ID delimiter

A one-or two-character symbol placed around IDs.

identifier

ID.

ID level

Part of an ID. For example, if the IDs of a text consist of page and line numbers, page numbers are one level and line numbers are another.

ID level delimiters

A one-or two-character symbol used to separate levels of an ID.

interactive

> Computing providing immediate response to user requests from a remote terminal.

job

> A task performed by a computer.

keyword

> A word in a text to be listed with its contexts in a concordance; opposed to stopword. Also, in a LEXICO command, a reserved word.

lemma

> Base.

lemmatization

> Headword classification.

list (v.)

> As used by LEXICO, to print information on the MACC printer.

LOOKUP

> The process of associating a base form with each entry in the word list of a text using the classifications stored in the basetype rules of a collection.

MACC

> Madison Academic Computing Center.

MACC tape number

> A number assigned by the MACC tape librarian to a tape.

mass storage

> On-line storage devices such as disks.

note delimiter

> A one-or two-character symbol placed before and after a note to separate it from the citation with which it is associated.

null string

    The string containing no characters.

numeric ID

    An ID in which every level is a number.

numeric ID level

    An Id level that is a number.

odd parity

    An error-checking technique used in writing and reading magnetic tapes.

off-line

    Referring to a computing task that is performed away from a remote terminal and whose results are printed at the central computer site. All off-line jobs at the computing center are placed in a pool and run according to user-assigned priorities. To initiate an off-line job, LEXICO asks the user to specify the desired priority; the user then receives a run identification.

offset

    A number added to all sequence numbers given in editing commands so that a user can refer to the sequence numbers that appear on a text listing even though citations have been inserted or deleted in previous EDIT blocks.

on-line

    In direct communication with a running computer program, via a remote terminal; providing immediate response to user commands and requests.

original form

    Text form; a word as it occurs in a text; one of the three forms of a type stored in the word list of a text.

pack

    To rearrange the contents of a file so that a minimal amount of space is required.

parameter

An entity such as word delimiters which can be assigned different values.

pass word

A code selected by the user which along with his user number identifies him to the UNIVAC 1110.

physical length

Total number of characters; for example, in a line.

physical tape block

A unit of readable information on magnetic tape.

pointer

A marker used in an EDIT block to indicate where in the text a change should be made.

priority

A job classification determining how soon a job will be completed and the rate at which it will be billed.

project number

An account number assigned by MACC to a project.

prompt

A request by LEXICO for the user to enter information.

qualifier

A prefix for a file name.

rejected type

A word that has been removed from the word list of a text.

remote terminal

A device used to communicate with a computer, usually consisting of at least a typewriter-like keyboard and connected to the computer by phone line.

removable characters

>Characters to be stripped off the beginning or end of a word before the word is entered into a concordance; frontstrip, endstrip, and squeeze out characters.

reserved character

>A character, such as a blank, comma, hyphen, colon, or digit, which cannot occur in information given to LEXICO within a command unless the string containing it is enclosed in single quotation marks.

reserved word

>A word, such as SHOW, DELIMITER, or ADD, with a specific use in LEXICO commands. To enter reserved words as information, enclose them in single quotes; for example, ADD STOPWORD 'ADD' ; .

RESPELL

>The process of normalizing the spellings of words in the word list of a text by application of spelling rules.

respelled form

>The form of a word after the application of spelling rules; one of the three forms of each word stored in the word list of a text.

reverse

>A declaration specifying that those types which would otherwise be treated as stopwords should be treated as keywords and vice versa.

run

>A task performed by a computer or the act of performing such a task.

run identification

>A name assigned by the system to an off-line job to identify the task and its output at the computing center.

run statement

>The first statement required of all jobs run at MACC. (See the MACC Computing Handbook.)

SDF file

    A file format unique to the UNIVAC 1110.

sequence numbers

    Numbers assigned by LEXICO sequentially (1, 2, 3, ...) to information
that can be referred to by number by the user; specifically
      (1)   the citations of a text ;
      (2)   the citations listed under a keyword in a concordance
      (3)   the types in the word list of a text; and
      (4)   the bases of a homograph.

single-statement block

    A command which completely specifies a task in one statement; it may not
be followed by an END statement.

single quote

    The symbol '(apostrophe).

slips

    Any alphabetized listing giving base forms and the contexts in which
their text forms occur; more specifically, a listing in which each page
contains a single citation from a text, one of its keywords, and the base
of that keyword.

spelling rule

    A direction to replace all occurrences of one specified string within the
words of a text by another specified string during the RESPELL process.

square brackets

    The symbols [and].

squeeze out

    Characters that appear in a text but are ignored in the formation of tokens.

statement

    Command.

stop character

    A character which, when appearing at the beginning of a word, designates that word as a stopword; autostop.

stoppable length

    A designation that all words of a certain length are to be treated as stopwords.

stopword

    A type that appears in a text but is not listed with its citations in a concordance; only its frequency is given.

string

    A finite sequence of characters.

system default

    A parameter value that will be used by LEXICO if no alternative is specified by the user.

task command

    A request to LEXICO to perform a process; a block header or single-statement block.

temporary text value

    A value which holds only during the operation (block) in which it was defined.

text

    A sequence of words and other character strings, arranged in citations, and delimited by a title at the beginning and a text delimiter at the end.

text code

    An identifying number assigned to a text by LEXICO when the text is entered into a collection.

text default

    A parameter value which holds for a text when no value is declared by a user.

text delimiter

A one-or two-character symbol used to mark the end of a text.

text form

Original form; a word as it appears in a text.

text name

A string of up to 12 characters used to identify a text.

text portion

The part of a citation that is concorded; the citation without its ID or notes.

time limit

A maximum limit on computer time set by a user for a job. If that limit is reached before the job is completed, the UNIVAC 1110 terminates job processing.

token

Any occurrence of a word in a text.

top

The beginning of a text--the point before the first citation (used in an EDIT block).

track

The number of recording bits on magnetic tape for a single character (including parity bits); generally 7 or 9.

type

A unique word in a text; opposed to token.

unmatched

An entry in the word list of a text that has not been assigned a base form.

user number

> A MACC-defined number for each user.

word

> A string of characters from a text entered into a concordance; keyword
> or stopword (see also type, token, frontstrip, endstrip).

word delimiter

> A one- or two-character symbol, such as a blank, that occurs between
> words of a text.

word list

> A list of all keywords and stopwords that appear in a text, created when
> the text is concorded. A respelled and a base form may be associated with
> each entry in a word list.