

LEXICO Guide No. 7

HEADWORD CLASSIFICATION

Richard L. Benezky
Nathan Relles
Lynne A. Price

Computer Sciences Technical Report #288

December, 1976

LEXICO Guide No. 7

HEADWORD CLASSIFICATION

Richard L. Venezky
Nathan Relles
Lynne A. Price

Computer Sciences Department
University of Wisconsin
December, 1976

TABLE OF CONTENTS

1. General Description	1
1.1 Introduction	1
1.2 Rules and Word Lists	1
1.3 Basic Processing	2
2. Respelling	4
2.1 Spelling Rules	4
2.2 RESPELL	5
3. Basetype Rules and LOOKUP	8
3.1 Basetype Rules	8
3.2 Displaying Basetype Rules	9
3.3 Homograph Rules	9
3.4 Copying Basetype Rules	11
3.5 LOOKUP	11
4. CLEANUP--Part 1: General Processing	13
4.1 The CLEANUP Block	13
4.2 Finding Unmatched Forms	13
4.3 Inspecting the Word List	14
4.4 Changing the Three Forms of a Word	15
4.5 Counting Entries in the Word List	16
4.6 Removing Entries From the Word List	16
4.7 Deleting Respelled Forms	17
4.8 Text Memo	17
5. CLEANUP--Part 2: Homographs	18
5.1 Homograph Bases	18
5.2 Assigning Citations to Homograph Bases	19
6. Slips	21
6.1 Three Output Options	21
6.2 Slips	22
6.3 Slips File	22
7. Listings	24
8. Abbreviations	26

TABLE OF CONTENTS cont.

9. Summary of Commands	27
9.1 Adding, Deleting, and Displaying Rules	27
9.2 CLEANUP	27
9.3 Blocks and Single-Statement Blocks	28
10. Reserved Words	29

1. General Description

1.1 Introduction

Concordance entries are listed alphabetically according to the spellings of the text forms, that is, according to the spellings of the words that appear in a text. However, for most users, a more desirable arrangement for a concordance is according to entry forms of a dictionary, or some other type of normal forms; for example, with present and past tense forms grouped under the same entry. LEXICO has a complete facility for aiding the user in making such classifications. This facility does require the user to designate an entry form (also called base, headword, or lemma) for each unique text spelling, and to indicate for homographs which citations in a concordance go with which entry forms, but several features of the system greatly reduce the repetitive work that would be done if the classification were done totally by hand. These features include storage and application of spelling rules for normalizing spelling; storage and application of word classification rules (here called basetype rules); on-line classification of homographs; on-line correction of unmatched and mismatched forms; and listing the concordance entries--text forms and citations--alphabetically by the spellings of the headwords.

1.2 Rules and Word Lists

Whenever, a text is concorded, a list of all the types (keywords and stopwords) which occur in the text is stored in the collection. This list is alphabetized and the words are numbered with the same sequence numbers printed with the concordance. Users may choose to associate with each original form on the list a respelled form and a base form. To do so, they request that spelling and basetype rules be applied to the word list. Of course, they must first

define the particular rules they wish to use. The latter is done by specifying rules in a CREATE or UPDATE block. Thus, one set of rules is associated with an entire collection; word lists, however, are text-specific. Any changes made to the rules of a collection do not affect a word list unless the new rules are explicitly applied to the list.

1.3 Basic Processing

If spellings are to be normalized, spelling rules are specified by the user (see Section 2.1). These rules are then applied with the RESPELL command to the entries in a word list before the words in this list are classified by base forms (see Section 2.2). For example, in Old English texts, *eth* and *th* can be converted to thorn so that when classification rules are specified, only the thorn spelling need be used.

For headword classification, LEXICO stores a list of basetype rules which can be applied to original or respelled forms. This file is normally built up slowly over time. An initial set of rules is added (see Sections 3.1 and 3.3); words from the first concordance are processed (called LOOKUP, see Section 3.5); the new rules are added for those words which were not matched and erroneous rules are corrected (altogether called CLEANUP, see Section 4). (Homographs are also divided during CLEANUP but this explanation is delayed until Section 5). This process of an initial LOOKUP followed by one or more user-directed, on-line CLEANUPS is then repeated for each succeeding text, with the number of unmatched forms in the LOOKUPS presumably growing gradually smaller.

Repeated requests for either RESPELL or LOOKUP on the same text will cause replacements of the forms generated by previous requests for the same processes.

CLEANUP blocks are used for correcting and adding base forms to a word list and are thus text-specific. However, when the user enters a base, LEXICO asks if the pairing just made should become a rule. Thus, changes made in this text-specific block may affect the rules of the entire collection. For the user's convenience, any of the commands which refer to the collection's spelling and basetype rules may be entered in a CLEANUP block. Usually, however, rules are entered in a CREATE or UPDATE block. Some users may prefer to perform initial matchings with CLEANUP rather than with LOOKUP. This may be especially convenient for the first text processed, since the user can avoid entering rules explicitly by adding them to a collection only when they are suggested by LEXICO.

CLEANUP blocks are also used for correcting errors in respelled forms, for assigning citations to homograph bases, and for removing unwanted words from the word list. There is no limit to the number of CLEANUP blocks used for a single text.

After the basetype pairings have been completed, the SLIPS command (see Section 6) may be used to print the concordance entries either as a concordance or as separate records (slips) for each citation. This statement is also used to obtain slip images on a disk file (which may be copied to tape).

2. Respelling

2.1 Spelling Rules

Spelling rules are added in a CREATE, UPDATE, or CLEANUP block with the command

```
ADD SPELLING RULES r1 : s11 s12 ... , r2 : s21 s22 ... ;
```

where r1, r2 etc. are the normal spellings and s11, s12, s21, etc. are text spellings. This means that when a RESPELL is done, all occurrences of any of s11, s12, ... s1n are replaced with r1, and so on. Either r or s may be up to five letters in length and s may include an initial or final word delimiter to indicate a rule which applies only to a word-initial or word-final spelling. (When a reserved character, such as a blank occurs in an s string, the entire string must be enclosed in single quotes.) If r is the null string (entered as two consecutive quotes) then all occurrences of s are deleted.

If conflicting rules are entered, an error message is given and the most recently entered rule is rejected.

Example 1:

```
ADD SPELLING RULE x : ts ;
(Replace all occurrences of ts with x.)
```

Example 2:

```
ADD SPELLING RULES * : th t ;
(Replace all occurrences of th and of t with *. Note that successive s strings are separated by blanks).
```

Example 3:

```
ADD SPELLING RULES b : 'bb ' , d : 'dd ' , f : 'ff ' ;
(Replace all geminate final b, d, and f with corresponding single letters. Note that when a blank occurs in an s string, the entire string is enclosed in quotes).
```

The command

```
DELETE SPELLING RULES r1 : s11 s12 ... , r2 : s21 s22 ... ;
```

can be used to delete previously entered rules. For example, the sequence


```
ADD SPELLING RULES rr : r1 rxx '-r' ;
DELETE SPELLING RULES rr : rxx '-r' ;
```

is equivalent to

```
ADD SPELLING RULE rr : r1 ;
```

There is a limit of 100 spelling rules which may exist in a collection at any one time.

To delete all spelling rules, enter

```
DELETE ALL SPELLING RULES ;
```

Spelling rules, once entered, are stored in the collection until deleted.

To review the current spelling rules on-line, enter

```
SHOW SPELLING RULES ;
```

An off-line listing of spelling rules may also be obtained (see Section 7).

2.2 RESPELL

RESPELL causes the current set of spelling rules to be applied to the word list of a specified text. Spelling rules are applied left to right, longest to shortest, and nonrecursively. This means, for example, that for the rules

```
p : x
aa : xx
z : aaa
```

the text form raxx is respelled raaa and not rapp or rz. For the rules

```
phi : rax
eh : ax
```

raxx is respelled phix and not rehx. For the rule

```
'' : ab
```

(where '' denotes the null string) aabb is respelled ab.

If during RESPELL a word disappears entirely, its respelled form is set to ??????. Thus, for the rules

```
'' : a
'' : bcd
```

the text form abcdaa is respelled ??????.

Once the desired spelling rules have been entered in a collection, the word list of any concorded text may be respelled. The process is initiated by the single statement block

```
RESPELL text ;
```

where text is the name or code of the text. An alternate command is

```
RESPELL ;
```

LEXICO responds to the shorter statement by prompting WHICH TEXT? Either the text name or the text code may be given.

The RESPELL command does not require an END statement. When it is entered, LEXICO first verifies that the text has been concorded. If not, no word list exists for the text, and an error message explains that the RESPELL cannot be performed.

If a word list does exist when the command is entered, LEXICO asks HERE AND NOW? An answer of y (for 'yes') indicates that the RESPELL should be performed at once, on-line. An answer of n (for 'no') indicates that the RESPELL should be performed off-line as a separate job.

LEXICO then prompts WANT A LISTING? If the response is y and the RESPELL is on-line, current spelling rules and all original forms with the corresponding respelled forms are displayed at the user's terminal. If the RESPELL is done off-line, the listing is printed at the computing center. If the response is n, the respelled forms are entered in the word list, but not displayed or printed.

If the RESPELL is to be done off-line, LEXICO asks the user if a backup of the collection should be created immediately before the RESPELL is performed (see Guide 3, Section 6). The system then prompts for a job priority. This affects how quickly the job is run and how much it costs. The options are explained in Guide 2, Section 7.

Spelling rules may be changed and RESPELL performed on a text as often as desired. RESPELL erases any base forms which may have been assigned in a word list.

To correct errors that result from RESPELL, such as exceptions to general rules, or to remove all respelled forms from a word list, use the CLEANUP block (see Section 4).

3. Basetype Rules and LOOKUP

3.1 Basetype Rules

Basetype rules are entered in a CREATE, UPDATE, or CLEANUP block. The form of the statement is

```
ADD BASETYPE RULES b1 : t11 t12 ... , b2 : t21 t22 ... ;
```

where b1, b2, etc. are base forms and t11, t12, t21, etc. are corresponding text (or respelled) forms. For example,

```
ADD BASETYPE RULES be : am is was are ;
```

is used to classify am, is, was, are and be as forms of be. Note that be is automatically classified as a form of itself. If be is to be classified under a base other than itself, this assumed classification must first be deleted.

When entering rules, the ordering of the t words is unimportant. If either a t or a b word contains blanks or other reserved characters, it must be enclosed in single quotes. If conflicting rules are added, an error message is given and the most recently entered rule is rejected.

Basetype classifications may be deleted by the command

```
DELETE BASETYPE RULES b1 : t11 t12 ... , b2 : t21 t22 ... ;
```

To delete a classification for a single type, the following command is available:

```
DELETE BASETYPE RULE FOR type ;
```

where type is an original (or respelled) form.

To delete all basetype rules for words with a specified first letter, enter

```
DELETE ALL BASETYPE RULES FOR letter ;
```

To delete all basetype rules, the similar command

```
DELETE ALL BASETYPE RULES ;
```

is used.

There are also commands which can be used to change part of a rule:

```
REPLACE TYPE oldtype WITH newtype ;
REPLACE BASE oldbase WITH newbase ;
```

Note that the first of these commands applies only to the one rule for type oldtype. However, the second command may change several rules since any number of types may have the same base.

3.2 Displaying Basetype Rules

In a CREATE, UPDATE, or CLEANUP block, on-line displays of rules may be obtained by the following commands. To display any basetype rule for the form type, enter

```
SHOW BASETYPE RULE FOR type ;
```

To display all basetype rules for forms with a specified first letter, enter

```
SHOW ALL BASETYPE RULES FOR letter ;
```

To display all basetype rules, enter

```
SHOW ALL BASETYPE RULES ;
```

The rules will be grouped by the first letter of the type, but will not be alphabetized within these groups.

Off-line listings of these rules may also be obtained (see Section 7).

3.3 Homograph Rules

Single text forms often must be classified under more than one base.

Consider, for example, the string wind in the sentence, ' I wind my watch when

the wind blows'. LEXICO allows the user to classify such homographs under multiple bases and to specify which occurrences of the text form belong with each base. The latter is done in CLEANUP (see Section 5). However, a special kind of basetype rule, called a homograph rule, may be used by LOOKUP to assign variant bases. These rules may be added and deleted with the following commands:

```
ADD HOMOGRAPH RULES t1 : b11 b12 ... , t2 : b21 b22 ... ;
DELETE HOMOGRAPH RULES t1 : b11 b12 ... , t2 : b21 b22 ... ;
```

where t1, t2, etc. are original (or respelled) forms and b11, b12, b21, etc. are base forms.

Note that text (or respelled) forms occur on the left of homograph rules but on the right of other basetype rules. There is a limit of seven bases which may be specified for any one type ; however, these may be entered with one or more ADD HOMOGRAPH RULE commands.

If a single classification exists for a type when homograph rules are added for it, the existing base is retained. Thus, the sequence

```
ADD BASETYPE RULE led : lead ;
ADD HOMOGRAPH RULE lead : lead(n) ;
```

is equivalent to the single command

```
ADD HOMOGRAPH RULES lead : led lead(n) ;
```

Homograph rules are treated as a kind of basetype rules--they are included whenever the latter are displayed or listed. Also, the commands

```
DELETE BASETYPE RULE FOR type ;
DELETE ALL BASETYPE RULES ;
DELETE ALL BASETYPE RULES FOR letter ;
REPLACE TYPE oldtype WITH newtype ;
REPLACE BASE oldbase WITH newbase ;
```

all apply to any homograph rules which exist for the specified types.

3.4 Copying Basetype Rules

Some users may have their texts arranged in several collections and yet wish to use the same basetype rules for all the texts. To aid in doing so, the command

```
COPY BASETYPE RULES FROM cname ;
```

may be used to copy all basetype and homograph rules from collection cname to the current collection. This will delete any previously defined rules which are not duplicated in cname. To minimize the effort involved in recalling which collection contains the most complete set of basetype rules, it is recommended that users process the texts in one collection as completely as possible before switching to another collection.

3.5 LOOKUP

Once the desired basetype and homograph rules have been entered in a collection, a LOOKUP may be performed on the word list of any concorded text. LOOKUP is initiated by the single statement block

```
LOOKUP text ;
```

where text is the name or code of the text. Alternatively, the command

```
LOOKUP ;
```

may be used. In this case, LEXICO prompts WHICH TEXT? Either the text name or text code may be given.

The LOOKUP command does not require an END statement and results in an off-line job. If the text has not been concorded when the command is entered, an error message explains that no LOOKUP can be started. If a word list does exist, LEXICO asks the user if a backup should be created immediately before

the LOOKUP is done (see Guide 3, Section 6). The system then asks for a job priority which affects how quickly the job is run and how much it costs. These options are explained in Guide 2, Section 7.

LOOKUP causes the respelled forms, if any exist, to be compared to the basetype rules, and base forms extracted for all matches. If no RESPELL has been done, matching is based on text forms. A listing of all text forms with their corresponding respelled forms (if any) and base forms (if a match was found) is printed.

4. CLEANUP--Part 1: General Processing

4.1 The CLEANUP Block

During the on-line process called CLEANUP, bases for unmatched forms can be entered, incorrect bases or respelled forms replaced, citations for homographs assigned to bases, and unwanted words removed from the word list. To perform a CLEANUP on a text whose name or text code is text, use the following block:

```

CLEANUP text ;
    command 1 ;
    command 2 ;
        .
        .
        .
END ;

```

An alternate block header is

```

CLEANUP ;

```

If this variation is used, LEXICO prompts WHICH TEXT? Either the text name or text code may be given in response.

If the text has not been concorded, it has no word list, and no CLEANUP can be done. In this case, an error message is given when the block header is entered. The user should enter another task command--no END statement is required.

4.2 Finding Unmatched Forms

Among the commands which may follow the block header are those which cause the display of words to which no base form has been assigned:

```

SHOW UNMATCHED ;
SHOW UNMATCHED m - n ;

```

The former command causes the first unmatched word to be displayed, followed by a LEXICO-generated request for a base. After the user responds, the cycle is repeated for each successive unmatched entry. The latter command results in the display of unmatched entries whose sequence numbers in the word list are between m and n. With both commands, one entry at a time is displayed.

Once an unmatched form is displayed, LEXICO requests a base for it. The user may respond with a base; with a semicolon to terminate display of unmatched forms; with an equals sign to indicate that the base is identical to the respelled form (or to the original form if no respelled forms are present); with an h to indicate a homograph; with an r to reject the type (see Section 4.6); or with a question mark to leave the type unmatched. If any of these symbols (equals sign, semicolon, question mark, r or h) is itself a base, it should be enclosed in single quotation marks to distinguish it from the codes used here. If h is entered, LEXICO requests base forms and then citation numbers to assign to each base (see Section 5). With the entry of either a single base or homograph bases, LEXICO also asks whether or not to save the matchings as general rules (basetype and homograph rules) for further LOOKUPS.

4.3 Inspecting the Word List

To review a portion of the current word list on line, enter

```
SHOW TYPES m - n ;
```

Each word with a sequence number between m and n is displayed with its sequence number and all its forms. Thus, if a RESPELL has been done, three fields are displayed for each entry: the original form, the respelled form and the base form.

If no spelling rules were applied, only the original and the base forms are displayed. In either case, for unmatched types, the base form field is blank. Note that if a LOOKUP was performed, this display is identical initially to the printed list produced by the LOOKUP process. For original (or respelled) forms which are defined as homographs, all bases which have been defined for that type are displayed. As citations are assigned to bases (see Section 5.2) the citation numbers assigned to each base are also displayed.

An off-line listing of the current word list may also be obtained (see Section 7).

4.4 Changing the Three Forms of a Word

To correct any form of a word, first select the entry with the command

```
TYPE n ;
```

where n is the word list sequence number. Any of its fields may then be changed by a sequence of the commands

```
NEW ORIGINAL newo ;
NEW RESPELLED newr ;
NEW BASE newb ;
NEW BASE UNMATCHED ;
```

where newo, newr, and newb are the corrected forms.

If the original form is changed and respelled forms are being used, the form is respelled according to the current spelling rules. If there are no spelling rules, but respelled forms exist (that is, the rules have been deleted since the last RESPELL), the respelled form is set identical to the original form. Whether or not respelled forms are used, the base is then set

according to the current basetype rules. Similarly, if the respelled form is changed, the base is looked-up in the basetype rules. However, the original form is not changed. If homograph bases are specified by the basetype rules, LEXICO asks the user to assign citations to the bases (see Section 5). In all cases, after the changes are made, the new forms are displayed.

In processing the NEW BASE newb command, LEXICO asks the user whether or not the new base should also become part of a basetype rule. The NEW BASE UNMATCHED command is used to erase an existing base form without replacing it.

4.5 Counting Entries in the Word List

It is possible to find how many entries in the word list are still not matched with bases. The request is

HOW MANY UNMATCHED ;

To find the total number of types in the word list, enter

HOW MANY TYPES ;

4.6 Removing Entries From the Word List

The user may wish to discard some words (for example, Roman numerals) that appear in a word list. To remove a type from the list, enter

REJECT TYPE n ;

where n is the word list sequence number. A rejected type is not included when the word list is displayed or printed, nor when slips are generated. To replace a rejected type, enter

RESTORE TYPE n ;

4.7 Deleting Respelled Forms

To erase all respelled forms without replacing them, enter

```
DELETE RESPALLED TYPES ;
```

This command should be used when, after examining the results of a RESPELL, the user decides he would prefer to work with the original forms. It results in the deletion of all respelled types and all base forms found during LOOKUP or entered during CLEANUP. If the user then decides to add new spelling rules, they may be entered in the current block, or in a CREATE or UPDATE block. To request that the rules be applied to generate new respelled forms, however, requires a new RESPELL command.

4.8 Text Memo

The command

```
MEMO 'note about the text';
```

may be entered in any text-specific block, where note about the text is a string of up to 100 characters. This memo will be displayed whenever the status of the text is displayed (see Guide 3, Section 4.2).

5. CLEANUP--Part 2: Homographs

5.1 Homograph Bases

LEXICO allows as many as seven bases to be associated with a single original (or respelled) form. This may be done during LOOKUP by applying homograph rules or on-line during CLEANUP using the commands described here. To enter homograph bases in the word list during CLEANUP, first select the word with the TYPE n command. Then enter

```
ADD HOMOGRAPH BASES base1 base2 ... ;
```

As each base is added, LEXICO will display it with the base sequence number it is assigned. This sequence number may be used in later commands wherever the base could occur. If the type has an existing single base (assigned either by LOOKUP or CLEANUP) when this command is entered, this base will remain as one of the homograph bases. If the type is a keyword, LEXICO will ask the user to assign citations to each base (see Section 5.2).

To delete bases, enter

```
DELETE HOMOGRAPH BASES b1 b2 ... ;
```

where each b_i is either a base or the sequence number of a base.

The following command may be used to change a homograph base and retain any citations which have been assigned to it:

```
REPLACE .HOMOGRAPH BASE b WITH newbase ;
```

where b is either the base or the sequence number of the base and newbase is the corrected form. This command is not equivalent to a DELETE HOMOGRAPH BASE command followed by an ADD HOMOGRAPH BASE command, since DELETE HOMOGRAPH BASE also deletes any citation assignments associated with the base. Do not confuse

this command with the similar-looking REPLACE BASE oldbase WITH newbase command used for correcting rules.

5.2 Assigning Citations to Homograph Bases

Citations are assigned to homograph bases during CLEANUP according to their sequence numbers under the keyword on the concordance listing. Since citations are not listed under stopwords during a concordance, there are no sequence numbers that may be assigned to the individual bases of a stopword. There are several statements in CLEANUP where LEXICO asks the user to assign citations to a base. This prompting occurs only for keywords.

When answering a LEXICO-generated request for citations, enter

```
c1 c2 c3 ... ;
```

where each ci is a citation number, an inclusive sequence of citation numbers (e.g., 6-12, 1-3), or the word REST to indicate all remaining unassigned citations. (To postpone making citation assignments, simply enter ;). To assign citations when the system has not requested them use the command

```
ADD TO BASE b CITATIONS c1 c2 c3 ... ;
```

where b is either a base or the sequence number of a base, and the ci's are as above.

Here are some examples:

```
ADD To BASE lead(verb) CITATIONS 2 5-8 11 15 17-21 ;
ADD TO BASE 2 CITATIONS REST ;
ADD TO BASE wind(verb) CITATIONS 12 18 REST ;
```

In the last example, the twelfth and eighteenth citations listed under the keyword, which presumably is wind, are switched from some other base to

wind(verb) and all remaining unassigned citations are also assigned to wind(verb).

To destroy previously entered citation assignments without replacing them, enter

```
CLEAR CITATIONS c1 c2 ... ;
```

where each c_i is either a citation number or a sequence of citation numbers.

Homographs are considered unmatched unless a base has been associated with every citation. If an unmatched homograph is encountered while processing the SHOW UNMATCHED command, LEXICO prompts the user for additional bases and for citations to assign to new and old bases.

6. Slips

6.1 Three Output Options

LEXICO will generate a listing of concordance entries alphabetized by the spelling of the base forms in any of three forms. These are:

1) Slips

Each 5½" x 6½" page contains a base form with one of its text forms and a single citation containing the text form. (The respelled form is not given). The collection and text names are printed at the bottom of the page.

2) Base Concordance

A listing similar to a concordance is printed. The entries are alphabetized by the spelling of the base form. Each word is listed with its base form, respelled form (if any), and text form.

3) Slips File

A user may wish to have the slips printed in a format other than those just described. LEXICO will produce a data file which the user can format with his own program (see Section 6.3).

If a word is still unmatched when slips are generated, its base form field is left blank in all types of slips output.

Stopwords are listed with their frequencies in all forms of slips output. If homograph bases have been entered for a stopword, an entry is made for each base. However, the frequency for the first base is set equal to the actual number of times the stopword occurred in the text and the frequency of all other bases is set to zero. This is done so that the sums of the frequencies of all stopwords and keywords will be the total number of tokens that occurred in the text (except for any types which were rejected).

6.2 SLIPS

To cause any of the above outputs to be produced, enter

SLIPS FOR text ;

where text is the name or code of the text. Alternatively, enter

SLIPS ;

If the shorter command is used, LEXICO asks WHICH TEXT? Either the text name or text code may be given.

The SLIPS command is an entire block and does not require an END statement. When it is entered, LEXICO checks that the text has been concorded and that a word list exists. If not, an error message is given and no slips are generated. If the text has been concorded, LEXICO asks the user which forms of output should be produced. If the concordance was stored on tape (see Guide 6, Section 3.5), the system prompts for the MACC tape number of the tape, and the number of files on that tape preceding the one containing the concordance. While processing this command, LEXICO also asks the user for a job priority.

6.3 Slips File

If the user requests the slips file output option from the SLIPS command, LEXICO will prompt for the name of the file on which to write this information. The user must create the file outside of LEXICO (see Guide 2, Section 8.2). The user is also responsible for deleting the file when it is no longer needed.

The slips file is a FORTRAN-readable file. The first character of each line indicates the contents of the line. The second character of each line is a colon. The file contains

	<u>First Characters</u>	<u>Rest of line</u>
1.	N:	collection name
2.	T:	text name
3.	B:	base form
4.	R:	respelled form
5.	O:	original form
6.	S:	frequency (this line appears for stopwords)
7.	K:	frequency (this line appears for keywords)
8.	C:	sequence number of citation within keyword (for cross-reference to the printed concordance)
9.	X:	first line of citation
10.	X:	second line of citation
		.
		.
11.	X:	last line of citation
		.
		.
12.	E:	(marks end of file)

Lines of type 1 and 2 appear only once, at the beginning of the file. Lines 3-11 (except 6) are repeated for each keyword. Within these groups, lines 8-11 are repeated for each citation within the keyword. Lines 3-6 are repeated for each stopword. Line 12 appears after all keywords and stopwords have been given.

7. Listings

At the user's option, an output listing of text and respelled forms is produced by RESPELL. If such a listing is printed, it includes the spelling rules used. LOOKUP always generates an output listing. CLEANUP does not print current forms, but such a listing can be obtained with the command

```
LIST TYPES FOR text ;
```

where text is either a text name or text code. An alternate form of this command is

```
LIST TYPES ;
```

When the short form is entered, LEXICO asks WHICH TEXT? Either the text name or text code may be given. This command is its own block and therefore does not require an END statement. It may not appear, however, in any block. When it is entered, if the text has not been concorded, an error message explains that the word list cannot be printed. If the text has been concorded, LEXICO asks the user to enter a job priority.

To produce a listing of the spelling rules currently defined in a collection, use the single-statement block.

```
LIST SPELLING RULES ;
```

The system will prompt for a job priority.

Two further listing requests (which are also single-statement blocks and which also require jobs priorities) are:

```
LIST ALL BASETYPE RULES ;
```

and

```
LIST ALL BASETYPE RULES FOR letter ;
```

The former command causes all of the basetype and homograph rules defined in a collection to be listed; the latter prints only those rules for original or respelled forms beginning with the specified letter. Basetype rules may be listed alphabetized according to the spelling of the base, alphabetized according to the spelling of the type, or grouped by the first letter of the type but otherwise in no special order. It is more expensive to produce the alphabetized lists than the unordered one. When one of these commands is entered, LEXICO asks what type of the alphabetization, if any, should be performed.

8. Abbreviations

Members of the following groups of expressions may be used interchangeably during headword classification:

BEFORE, BY, EXCEPT, FOR, FROM, INTO, OF, ON, OR, OUT,
THRU, TO, WITH ;
ADD BASETYPE RULES, ABTR ;
ADD HOMOGRAPH BASES, AHB ;
ADD HOMOGRAPH RULES, AHR ;
ADD SPELLING RULES, ASR ;
BASE, BASES ;
BASETYPE RULES, BTR ;
CITATION, CITATIONS, CIT ;
DELETE, D ;
DELETE BASETYPE RULES, DBTR ;
DELETE HOMOGRAPH BASES, DHB ;
DELETE HOMOGRAPH RULES, DHR ;
DELETE SPELLING RULES, DSR ;
HOMOGRAPH, HOMO ;
HOMOGRAPH RULES, HR ;
REPLACE BASE, RB ;
REPLACE TYPE, RT ;
RULE, RULES ;
SHOW, S, SH, DISPLAY ;
SPELLING RULES, SR ;
TYPE, TYPES ;

9. Summary of Commands

9.1 Adding, Deleting, and Displaying Rules

The following commands may be entered in CREATE, UPDATE, or CLEANUP

blocks:

```
ADD SPELLING RULES r1 : s11 s12 ... , r2 : s21 s22 ... ;
DELETE SPELLING RULES r1 : s11 s12 ... , r2 : s21 s22 ... ;
DELETE ALL SPELLING RULES ;
SHOW SPELLING RULES ;
```

```
ADD BASETYPE RULES b1 : t11 t12 ... , b2 : t21 t22 ... ;
DELETE BASETYPE RULES b1 : t11 t12 ... , b2 : t21 t22 ... ;
DELETE BASETYPE RULE FOR type ;
DELETE ALL BASETYPE RULES FOR letter ;
DELETE ALL BASETYPE RULES ;
```

```
REPLACE TYPE oldtype WITH newtype ;
REPLACE BASE oldbase WITH newbase ;
```

```
SHOW BASETYPE RULES FOR type ;
SHOW ALL BASETYPE RULES FOR letter ;
SHOW ALL BASETYPE RULES ;
```

```
ADD HOMOGRAPH RULES t1 : b11 b12 ... , t2 : b21 b22 ... ;
DELETE HOMOGRAPH RULES t1 : b11 b12 ... , t2 : b21 b22 ... ;
```

```
COPY BASETYPE RULES FROM cname ;
```

9.2 CLEANUP

The following commands may be entered in a CLEANUP block:

```
SHOW UNMATCHED ;
SHOW UNMATCHED m - n ;
SHOW TYPES m - n ;
HOW MANY UNMATCHED ;
HOW MANY TYPES ;
REJECT TYPE n ;
RESTORE TYPE n ;
DELETE RESPELLED TYPES ;
MEMO 'note about the text' ;
```

Also, after selecting an entry in the word list with the command

```
TYPE n ;
```

any of the following may be entered:

```
NEW ORIGINAL newo ;
NEW RESPELLED newr ;
NEW BASE newb ;
NEW BASE UNMATCHED ;
ADD HOMOGRAPH BASES base1 base2 ... ;
DELETE HOMOGRAPH BASES b1 b2 ... ;

REPLACE HOMOGRAPH BASE b WITH newbase ;

ADD TO BASE b CITATIONS c1 c2 ... ;

CLEAR CITATIONS c1 c2 ... ;
```

9.3 Blocks and Single-Statement Blocks

The following task commands pertain to headword classification:

```
RESPELL text ;
LOOKUP text ;
CLEANUP text ;
SLIPS FOR text ;
LIST TYPES FOR text ;
LIST SPELLING RULES ;
LIST ALL BASETYPE RULES ;
LIST ALL BASETYPE RULES FOR letter ;
```


10. Reserved Words

The following words are reserved in a CLEANUP block:

ABTR	HOMO	TYPE
ADD	HOMOGRAPH	TYPES
ADDCONCORD	HOW	UNMATCHED
AHB	HR	UPDATE
AHR	IGNORE	WITH
ALL	INTO	
ASR	LIST	
AUTOSTOP	LISTS	
BACKUP	LOOKUP	
BASE	MANY	
BASES	MEMO	
BASETYPE	MINUS	
BEFORE	NEW	
BLANK	NO	
BTR	NOT	
BY	OF	
CHARACTER	ON	
CHARACTERS	OPTION	
CIT	OR	
CITATION	ORIGINAL	
CITATIONS	OUT	
CLEAR	PACK	
CLEANUP	RB	
COLLECTION	REJECT	
COLON	RENAME	
COMMA	REPLACE	
CONCORD	RESPELL	
CONCORDANCE	RESPELLED	
CONCORDANCES	REST	
COPY	RESTORE	
CREATE	RHB	
D	ROUTE	
DBTR	RT	
DELETE	RULE	
DHB	RULES	
DHR	S	
DIR	SEMI	
DIRECTORY	SEQUENCE	
DISPLAY	SH	
DO	SHOW	
DONOT	SLIPS	
DSR	SPECS	
EDIT	SPELLING	
END	STATUS	
EXCEPT	TEXT	
FOR	TEXTS	
FROM	THRU	
FREQUENCY	TO	