27. MacLex limits.

Most of the limits pertain to reversals. Here are the maximums which you should not exceed.

Total number of records in the lexicon   no limit
Total number of files in the lexicon   no limit
Total number of fields in an input record 300 (effective only for
                                                reversals, otherwise no
                                                limit)
Maximum number of fields in a reversed record no limit
Maximum number of instruction sets   5
Maximum number of instructions per set   20
Maximum number of subrecord header tokens   10
Maximum number of group header tokens   30
Maximum number of fields in a multifield set   6
Maximum number of subrecords in a record   40
Maximum number of groups in a record   60
Maximum number of characters in a token   7 (more, for lexicon mode
                                                or sorting, but it is wise
                                                to limit yourself to 7)
Maximum number of characters in a static field   255
Maximum number of characters in other field types   no limit
For Find and Change:
Maximum number of fields nominated for
searching in, or for not searching in   32
Maximum number of trace list entries   equals number of records
                                                when lexicon was loaded.

For Exporting:
Maximum number of fields nominated for
exporting, or not for exporting   32
Maximum number of fields nominated for searching
for a character to be removed or replaced   10
Maximum number of unique characters that can
be removed   3
Maximum number of unique characters that can
be replaced   3

For (Semantic) Categories, or other hierarchical table data.
Maximum number of first level categories   200
Maximum (total) number of second level categories   1000
Maximum number of third level categories   no limit
        (actually, the memory required for the 3rd level's list to be
        displayed must not exceed 32kb.)

MacLex is distributed with a recommended MultiFinder partition size of 560 kb. This is about as small as you should go for a 5000 record lexicon; it would allow you only about a dozen files in memory at any one time, and lots of file i/o might need to be done for that reason. You can increase the size by selecting the MacLex icon and choosing the Get Info item on the File menu at the desktop, then change the number in the box at bottom right to something higher than 560 - about 750 is a good compromise. MacLex will let you set the number as low as 368 kb. Lower than that you will probably have difficulty handling anything but a lexicon about a thousand entries or less.

MacLex (from version 1.1d onwards) has improved performance for tight memory situations. It now will not crash if the index table can't be built, or if it gets built but there is not enough memory to input lexicon files. In these situations MacLex will warn you, and then either quit, or allow you to continue - depending on how serious is the condition. In some cicumstances it will warn you to increase the MultiFinder partition size - requiring you to quit from MacLex first, of course. You can keep on editing if you wish - you'll get away with it for a while, but if MacLex crashes it will be your fault, not mine.

If you want to work out MacLex's memory requirements, the following are the facts you need to know.

The index table in memory has as many entries as records in the lexicon at load time, + 400 empty entries (for your additions).
Each entry has the following number of bytes:
      sort key length (default 16) + 2 + sort key length + 2 + 12
      and a further 10 bytes for file number, record offset,
      and record length values.
Thus a record entry, for the default setting, is 58 bytes long.
A 5000 record lexicon will need a table space of 5400 x 58 = 313200 bytes.

MacLex also keeps a space for the trace list entries: it is 2 x the number of records; for the above example that is 5000 x 2 = 10000 bytes.

The space required for storing semantic category index values is about 3.2kb. The semantic category table itself, if used, will reside in memory. The one supplied with MacLax takes about 18kb.

Your MultiFinder partition therefore has to fit these two tables in it, plus the MacLex application itself, the semantic category index tables and the category table, and still have at least 50kb for lexicon files (preferably much more - say, 150 kb for efficiency).

For file i/o as you work with your lexicon, MacLex wants enough free memory for the largest input file to fit in memory twice over, plus a bit of extra space for editing changes - if the extra space is less than 5kb, you will get a warning message and MacLex will quit; if less than 20kb, you will just be warned. These warnings only occur when you first load the lexicon.

28.   Other functions planned.

The following are major things I intend to add to MacLex as time permits. The easier ones will appear in early versions, harder ones in later versions.

28.1 Scanning a text file for unentered lexical items.

Implement scanning a text file, checking words intelligently against what is currently in the lexicon. Allowing user to easily identify and construct new lexicon records interactively.

28.2 Merge

The ability to merge one or more files of unsorted lexicon data, provided the records to be merged have the same header type as the receiving lexicon's header fields. I expect to build in various options such as unconditional merging, prompted merging when there is a matching header already (with the option to merge, or append as a subrecord, or skip), or prompted merging for every record. I want to also include a trace function which will allow the user after the merging is done to trace through the lexicon going only to those records inserted or changed by the merger process for each click of a button.

28.3 Token checking

An option to allow each typed token to be checked against a standard list. Variation are flagged giving the user a chance to either edit a wrongly typed token, or to add a new one to the list.

28.4 Combining and Uncombining records

A facility to allow the user to take a displayed record and make it a subrecord of a previously tagged record. Also the opposite capability, to take a portion of a record and make it into a record.

28.5 Formatted display & print options

Allowing MacLex to handle multiple column formatted printing with multiple fonts and sizes. Also to see records in formatted form on the screen (but not in multiple columns).