

Sort Subprogram Help

(Need another help file? Try [Macrex Help Contents](#). Macrex help keys <CTRL><ALT>F1 or <CTRL><ALT>F10)

[Sort subprogram main menu](#)

[Macrex Sort Options Menu](#)

[The sort definition file](#)

[Sorting batches](#)

[Alphabetical sequence used by MACREX](#)

[Sorting after merging backup files](#)

[Redefining the sort](#)

[How the sort works](#)

[Words ignored only at the beginning of a subheading](#)

[Sorting defaults](#)

[MACREX disk sort](#)

Usually MACREX sorts all the entries immediately ([Option J - Autosort](#) is set to **On** on [General options Menu 1](#)), so there will be no need to access the sort subprogram explicitly unless you want to change the way the sort works, or unless you want to re-sort an index that has previously been sorted using a different system. You might also need it if the index is too big to be handled with Autosort On, but this is unlikely because most computers can accommodate at least a hundred thousand entries with Autosort **On**. You can access the Sort Subprogram from the main menu by Options->Sort with options. There are four choices:

MACREX SORT SUBPROGRAM

24-12-2007, 13:28:37

[C](#) - Change sort options

[S](#) - Sort index

[H](#) - Sort Help

[M](#) - Main menu

Sorting the index

If you have changed the sort options and want to resort your index using the new options, press **S**.

If Autosort is **Off** and you want to [fine sort your index](#), press **S** from this menu.

Press **H** to display this help menu.

Press **M** to return to the [Main Menu](#).

If you wish to change the sort defaults press **C** to take you to the Sort Options Menu

MACREX SORT OPTIONS MENU

[A](#) - First letter sorted A

[B](#) - Last letter sorted Z

[C](#) - Drive for temporary files C

[D](#) - Merge following sort? Yes

[S](#) - Sort "see also" to end as ZZZ

[T](#) - Hyphen Fix for Chicago-15 No

E - Sort prepositions etc? No
F - Allow for text in curly brackets? Yes
G - Alphabetization - word by word
H - Sort numbers within entry? Yes
I - Maximum length of sorted numbers 10
J - Unconditional re-sort Yes
K - Edit list of ignored words
L - Edit list of changed words
M - Edit character translation table
N - "See also" references at end of list? Yes
O - "See also" text ^see also^
P - Sort roman Numbers within entry? No
Q - Replace soft commas? No
R - Soft comma replacement text >

<ESC> - Save current settings, ^L, ^S/F8 - Load or Save sort definition
 Change options as needed then press <return> ==>

The options A to J of this menu represent each feature and its current (default) state. You may want to change D to J quite often.

The next three options (K to M) allow you to define your own sorting principles – you will only want to change these if you are trying to do something unusual or are indexing in a foreign language for which we have not supplied a sort definition file. Options N to O deal with the merging of cross-references and P is concerned with roman numerals in entries. Options Q and R concern the placement of [soft commas](#) in the sort order. Option S sets the sorting position for *see also* references when you want them to be at the end of the list of subheadings and option T provides for the correct treatment of hyphens when the Chicago-15 sorting convention is used.

A - First letter sorted

B - Last letter sorted

These two options will only be used if you are not using autosort because you have a massive index. They allow you to sort only one alphabetical section of the index, while leaving another part alone. This may be useful to save time re-sorting a large index if only one or two changes have been made. In fact the index is sorted from the batch containing the first letter specified to the batch containing the last one. You can choose the same letter for the first and last to be sorted and thus sort only one batch.

C - Drive for temporary files

This will only be used if you are not using autosort and you are short of disk space on the drive containing the index. It is used to choose the drive on which you have the most spare space for temporary files. Normally it doesn't matter what you choose because MACREX doesn't need to make temporary files unless there are more entries in one batch than will fit in memory at one time.

D - Merge following sort?

This makes it possible for an index to be sorted and merged in one sequence (exactly as in option C on the Main Menu - Combined sort and merge) with the current default values for the MERGE subprogram. See the [Merge Options](#) for how to set these defaults.

E - Sort prepositions, etc?

The prepositions, conjunctions and articles defined in [Option K](#) (below) will be ignored for sorting purposes if **No** is selected but will be treated like any other word in the index (i.e. sorted) if **Yes** is selected.

F - Allow for text in curly brackets?

Answering No causes all text enclosed in { } to be ignored by the sort in all circumstances.

Answering Yes causes text enclosed in curly brackets to be treated as though it were moved to the end of the word it precedes - thus it can affect the sequence if the words are the same but the prefixes different. It may be useful to answer yes to the above question for indexes containing chemical names such as

```
{alpha-}aminobutyric acid
{beta-}aminobutyric acid
{gamma-}aminobutyric acid
```

in order to get the prefixes in the right order - all entries would be filed under aminobutyric acid.

G - Alphabetization

This option allows you to choose word-by-word or letter-by-letter sorting. Choosing word-by-word means that the space at the end of each word breaks up the sorting sequence - when two words begin with the same sequence of letters, the shorter word precedes the longer; in letter-by-letter sorting spaces up to the first comma are ignored and the sort reads the text as a continuous string of letters.

The following examples are taken from the British Standard BS3700:1988:-

Example of sorting: appearance of unsorted entries

```
^BAIE magazine^
Bagshaw{,} Malcolm A.
B.B. King plc
BBC ^see^ British Broadcasting Corporation
^Back to books^
Bag of bricks
Bagby{,} George
^Backbench diaries of Richard Crossman^
Beach{,} Edward Latimer
BSI ^see^ British Standards Institution
banking
Ba Khin{,} Sayagi U.
B and T cell tumours
^Be my Valentine^
B.E.A. ^see^ British European Airways
B Ceram R A ^see^ British Ceramic Research Association
Bank of England
bell-ringing
^BS3700^~!~
BAA ^see^ British Airports Authority
```

word-by-word sorting

```
B and T cell tumours
B Ceram R A see British Ceramic Research Association
Ba Khin, Sayagi U.
BAA see British Airports Authority
Back to books
Backbench diaries of Richard Crossman
Bag of bricks
Bagby, George
Bagshaw, Malcolm A.
BAIE magazine
Bank of England
banking
B.B. King plc
```

BBC see British Broadcasting Corporation
 Be my Valentine
 B.E.A. see British European Airways
 Beach, Edward Latimer
 bell-ringing

BS3700 BSI see British Standards Institution

letter-by-letter sorting

BAA see British Airports Authority
 Backbench diaries of Richard Crossman
 Back to books
 Bagby, George
 Bag of bricks
 Bagshaw, Malcolm A.
 BAIE magazine
 Ba Khin, Sayagi U.
 B and T cell tumours
 banking
 Bank of England
 BBC see British Broadcasting Corporation
 B.B. King plc
 B Ceram R A see British Ceramic Research Association
 B.E.A. see British European Airways
 Beach, Edward Latimer
 bell-ringing
 Be my Valentine
 BS3700
 BSI see British Standards Institution

A problem unique to proper names: *soft* commas are not noticed by the letter-by-letter sort, e.g.

Hill{,} Y.

will come AFTER

Hillyer{,} W.

If you don't want this to happen either use word-by-word sorting or use ~Hill~

H - Sort numbers within entry?

I - Maximum length of sorted numbers

If option **H** is set to **Yes** (the default setting) numbers occurring within the body of an entry (rather than page numbers) will be sorted as though they were numbers rather than text. For example

Factor 1~!~
 Factor 8~!~
 Factor 14~!~

If this option is set to No you will get the following sequence

Factor 1~!~
 Factor 14~!~
 Factor 8~!~

(The blocker (~!~ - F7) is necessary if the number immediately precedes the page numbers, otherwise it would be mistaken for a page number.)

Option I allows you to set the maximum length (in the number of digits) of numbers that can be recognized as such by Option H.

J - Unconditional re-sort

This should normally be set to **Yes** to ensure that the whole index is re-sorted when you press **S** from the [Sort Program Submenu](#), taking into account any changes to the sort that you have made. If you have autosort **off** because you have a very large index, you may wish to set this to **No**. In this case MACREX remembers how much of each batch is already sorted and sorts only as much as is necessary.

K - Edit list of ignored words

It is a common practice to ignore certain small words when sorting, for example and, as, with. This option allows you to define your own list of words that you want to be ignored by the sort. It applies only to words within the body of the text, not to the first word of an entry. If you wish to have the first word of an entry ignored you should put it in curly brackets as in {The }Times., etc. You can always make sure that any particular word is ignored even if it is not in the list, by placing it in curly brackets. However, if the word occurs frequently it will be easier to have it ignored automatically.

Selecting Option K from the MACREX Sort Options Menu displays the current list of ignored words beneath which is a menu.

LIST OF WORDS OR PHRASES TO BE IGNORED BY THE SORT

```

0   or
1   and
2   as
3   for
4   in
5   the
6   of
7   on
8   a
9   with
10

```

```

A - Add word to list
E - Edit or delete word from list

```

Press <return> when done ==>

You can add a word to the list, or edit or delete an existing word by selecting **A** or **E** respectively. The ignored words are displayed in inverse video so that you can check that the spaces are present. This is important because it is essential to flank an ignored word with spaces for the sort to work correctly. The sort program will scan all entries for each ignored word and remove any it finds before comparing the entry with the next entry. All ignored words are removed from the entry and replaced with a space. Obviously you don't want the ignored words to be permanently removed from your index, so MACREX works on a copy of the entry concerned (the *surrogate entry*). The comparison is actually done using the surrogate entry rather than the entry itself. The reason that ignored words must be flanked by spaces is that they would otherwise be detected accidentally as parts of other words. For example if the three letters "and" were detected on their own then the word "chandelier" would be converted to "chelier" in the surrogate entry and would obviously not be sorted correctly.

You can have up to 64 ignored words on the list, but when adding them remember the following three points.

- i. The words are removed from the entry in the order in which they are defined. If, for example, you wished to have the phrase "in the treatment of" ignored by the sort you would have to place it in the list above the words "in", "the", and "of".
- ii. Scanning for and removing ignored words is the most time-consuming part of the sorting process. If you define a huge number of ignored words it might slow the program slightly.
- iii. If you wish to have words ignored only at the beginning of a subheading or sub-subheading you should delete them from the list of ignored words and insert them in the list of changed words - [option L](#).

L - Edit list of changed words

The character-translation and ignored-words facilities are not themselves adequate for sorting indexes - phrases like *see also* are not ignored, neither are they sorted in alphabetical order. *See also* references are usually placed at the beginning or the end of the list of subheadings. It is also very common for abbreviations such as *St* to be sorted as the full word. The system of replaced words makes it possible to have such words and phrases sorted as you wish. When the surrogate entry is made it is first scanned for the list of words to be replaced. If any words in the list are found they are replaced by the alternative specified. You can see and modify the list of replaced words by using this option. You can delete words from this list or add new words or phrases up to a total of 32 replacements. The same minor cautions apply to replacements as to [ignored words](#). Some words will need spaces added at one end or the other in the original form (for example *St* as a short form of *Saint* would lead *Street* to be converted to *Saintreet* if it didn't have the trailing space) and some will not (for example *æ* should be converted to *AE* for sorting even if it occurs in the middle of a word).

LIST OF WORDS OR PHRASES TO BE REPLACED BY THE SORT

1 Mc	2 Mac
3 St	4 Saint
5 St.	6 Saint
7 ^see^	8 ;
9 æ	10 AE
11 Æ	12 AE

A - Add replacement to list

E - Edit or delete replacement from list

Press <ENTER> when done ==>

- To **add a replacement** press **A** and follow the prompts.
- To **edit a replacement** press **E**, enter the number alongside the text to be replaced, edit it if required, press <ENTER> and then edit the replacement form if required and then press <ENTER> again.
- To **delete a replacement** press **E**, enter the number alongside the text to be replaced, backspace to clear the text and then press <ENTER>.
- The list of changed words is stored in the sort definition file ([see below](#))

M - Edit character translation table

This character translation table ensures that each individual character is sorted in the right place. For example it makes sure that upper case characters and lower case characters are treated in the same way by the sort. It will also ensure that accented characters are sorted in line with their non-accented equivalents. It does this by "translating" each character into an equivalent one for the sort (for example by default MACREX sorts **ð** - *eth* - as **D**). You will not normally need to change this table unless you are indexing using words from a language that uses letters additional to the 26 in

the A-Z alphabet that are not already accommodated by MACREX.

Choosing this option will result first in your being asked whether you want the word-by-word table, the letter-by-letter table or the first letter table. MACREX keeps independent copies of each available - they are selected by the G option above. When you have chosen the table all the ASCII codes from 0 to 127 are displayed along with the characters to which they correspond and those to which they are translated for sorting purposes.

Insert ON 0

Press W to edit word-by-word table, L for letter-by-letter table,

F for first letter table ==> W

0	◇ ◇	16	◇ ◇	32		48	0 A	64	@ @	80	P p	96	` ◇	112	p p
1	◇ ◇	17	◇ ◇	33	! ◇	49	1 B	65	A a	81	Q q	97	a a	113	q q
2	◇ ◇	18	◇ ◇	34	" ◇	50	2 C	66	B b	82	R r	98	b b	114	r r
3	◇ ◇	19	◇ ◇	35	# #	51	3 D	67	C c	83	S s	99	c c	115	s s
4	◇ ◇	20	◇ ◇	36	\$ \$	52	4 E	68	D d	84	T t	100	d d	116	t t
5	◇ ◇	21	◇ ◇	37	% %	53	5 F	69	E e	85	U u	101	e e	117	u u
6	◇ ◇	22	◇ ◇	38	& &	54	6 G	70	F f	86	V v	102	f f	118	v v
7	◇ ◇	23	◇ ◇	39	' ◇	55	7 H	71	G g	87	W w	103	g g	119	w w
8	◇ ◇	24	◇ ◇	40	(◇	56	8 I	72	H h	88	X x	104	h h	120	x x
9	◇ ◇	25	◇ ◇	41) ◇	57	9 J	73	I i	89	Y y	105	i i	121	y y
10	◇ ◇	26	◇ ◇	42	* *	58	:	74	J j	90	Z z	106	j j	122	z z
11	◇ ◇	27	◇ ◇	43	+ +	59	;	75	K k	91	[◇	107	k k	123	{ ◇
12	◇ ◇	28	◇ ◇	44	, =	60	< ◇	76	L l	92	\ ◇	108	l l	124	
13	◇ ◇	29	◇ ◇	45	- ?	61	= =	77	M m	93] ◇	109	m m	125	} ◇
14	◇ ◇	30	◇ ◇	46	. ◇	62	> ◇	78	N n	94	^ ◇	110	n n	126	~ ◇
15	◇ ◇	31	◇ ◇	47	/ @	63	? ?	79	O o	95	_ _	111	o o	127	◇

Enter number to edit, P to toggle page, <ENTER> when done ==>

The screen is only shows 128 of the 256 possible characters at a time. You can see the other half by typing P<ENTER> at this prompt - and get back to this screen by doing the same again.

Insert ON 0

Press W to edit word-by-word table, L for letter-by-letter table,

F for first letter table ==> W

128	Ç c	144	É e	160	á a	176	° °	192	À À	208	ð d	224	ó o	240	
129	ü u	145	æ e	161	í i	177	◇ ◇	193	Á Á	209	Ð d	225	ß b	241	± ±
130	é e	146	Æ e	162	ó o	178	² ²	194	Â Â	210	Ê e	226	Ï o	242	◇ ◇
131	â a	147	ô o	163	ú u	179	¡ ¡	195	Ã Ã	211	Ë e	227	Ò o	243	¾ ¾
132	ä a	148	ö o	164	ñ n	180	◇ ◇	196	Ä Ä	212	È e	228	Ö o	244	¶ ¶
133	à a	149	ò o	165	Ñ n	181	Á a	197	Å Å	213	Ï i	229	Õ o	245	§ §
134	ã a	150	û u	166	ª a	182	Â a	198	ä a	214	Í i	230	µ m	246	ö ö
135	ç c	151	ù u	167	º o	183	À a	199	Å a	215	Î i	231	þ z	247	÷ ÷
136	ê e	152	ÿ y	168	¿ ¿	184	© ©	200	È È	216	Ï i	232	þ z	248	° °
137	ë e	153	Ï o	169	® ®	185	¹ ¹	201	É É	217	Ù Ù	233	Ú u	249	… …
138	è e	154	Ü u	170	¬ ¬	186	º o	202	Ê Ê	218	Ú ú	234	Û u	250	. .
139	ì i	155	ø o	171	½ ½	187	» »	203	Ë Ë	219	Û Û	235	Ü ü	251	¹ ¹
140	î i	156	£ £	172	¼ ¼	188	¼ ¼	204	Ì Ì	220	Ü Ü	236	Ý ý	252	³ ³
141	ï i	157	Ø o	173	í i	189	¢ c	205	Í Í	221	Ý Ý	237	Ý ý	253	² ²
142	Ä a	158	× ×	174	« ◇	190	¥ ¥	206	Î Î	222	Ï I	238	î î	254	◇ ◇
143	Å a	159	ÿ ÿ	175	» ◇	191	¿ ¿	207	ß ß	223	ß ß	239	ï ï	255	◇ ◇

Enter number to edit, P to toggle page, <ENTER> when done ==>

Each screen consists of eight sets of three columns. The first item in each column is the ASCII (numerical) code for the character. The second is the symbol that this code will produce on the screen using the current [code page](#). The third is the character that you want to use when the sort program compares the entries. In the tables above the symbol ◇ is used to denote a character that will be ignored altogether by the sort. If you want to have a character ignored by the sort enter that you want is converted to for the sort as an ASCII code of 1 ([see below](#)). Note that the characters in the sample table shown above may not be identical to those that appear on your screen.

If you want to change the translation of one of the characters enter its ASCII code. You will then be invited to enter the character to which you want it converted for the sort. You can enter this in one

of two ways:

- i. Enter the ASCII code of the character you want.
- ii. Enter the actual character you want preceded by an closing single quote, '. If the character you want is a closing single quote simply put ' ' (two single quotes, not a double quote).

The options K, L and M options may be used to redefine the sorting sequence in almost any way - [see below](#) for an example of a sorting order for Icelandic.

N - "See also" references at end of list?

All entries are scanned for the words ^see also^, or for the text you have entered if you have changed (and saved) the default text in the MERGE subprogram. If ^see also^ or the equivalent text is found and this option is set to **Yes** then the entry containing it is moved to the end of the list of subheadings. If this option is set to **No** then it is moved to the beginning. If you chooses to have it sorted to the end of the list of subheadings the [text that you have chosen](#) for "see also" references will be converted to the text defined in [option S](#) (normally zzz) below for the purpose of sorting. For Danish and Swedish sorting orders see [option S](#) below.

O - "see also" text

This simply serves to remind you of the current text being used for "see also". If you wish to change it you must do so from the MERGE subprogram [Option R - See also text](#).

P - Sort roman numbers within an entry?

If this is set to **No** then roman numerals within the body of an entry will be sorted alphabetically, for example

```
Louis IX~!~
Louis VII~!~
Louis VIII~!~
Louis X~!~
```

while if it is set to **Yes** they will be sorted in numerical order

```
Louis VII~!~
Louis VIII~!~
Louis IX~!~
Louis X~!~
```

Note that in this example the text ~!~ is added to the end of the entry as a *blocker* so that MACREX does not mistake the roman numerals for roman page references. F7 has been assigned to produce this *blocker sequence*. Note also that if this option is turned on it may cause mis-sorting of some words which are both roman numerals and legitimate words or initials (for example I and mix and Jones DC). MACREX will only assume that it has found a roman numeral if it is a discrete word which is a valid roman number so it is not likely that problems will arise very often from the use of this option. However if it does cause problems you can switch the option off and used the keyword technique using the NUMBER program.

Roman numerals within an entry are not recognized if they appear as the first character of an entry. If this is needed for any reason you will need to use tildes and curly brackets.

See also General options [H- Recognize roman numbers](#).

Q - Replace soft commas?

R - Soft comma replacement text


```

black, hues of 89
black, Panatone 44
Black{,} Anna 34
Black{,} Zoe 22
Black Hawk Ale 55

```

The above is usually the required sequence that occurs when the same word is used as a proper name a noun and a adjective. The default setting for MACREX word-by-word sort will produce the following (incorrect) order

```

black, hues of 89
Panatone 44
Black, Anna 34
Black Hawk Ale 55
Black, Zoe 22

```

If **Q** is set to **Yes** and **R** is set to > (word by word) or @ (letter by letter) the correct sorting sequence will be achieved. This is because Setting **Q** to **Yes** makes the sort replace all the soft commas by the replacement text in the [surrogate string](#) before the comparison. On the Word-by-Word sort a hard comma is replaced by = and = is not translated in the translation table, and > is one above = in the ASCII sequence so the entry is sorted after the subheadings. On the Letter-by-Letter sort the default translation for a hard comma is ? (the next character before @) and the soft comma is (by default) replaced with @. If you wanted the soft comma entries to precede (perish the thought!) the subheadings you could change the soft comma replacements to < (wbw) and > (lbl) respectively.

Notes:

- i. This relies on angle brackets <> not being translated in the character translation table. If you are using these in the index and want them ignored and problems such as the above do not occur, you can change their translation to [ASCII 1 to have them ignored](#). Otherwise you can put the angle brackets in curly brackets.
- ii. Option **Q** (soft comma replacement text) is automatically changed from > to @ and back when you change the sort from Word-by-word to Letter-by-letter and back.
- iii. If you load a previous sort definition file, depending when it was made, it may change the translation of a hard comma and <> resulting in incorrect sorting if **Q** is set to **Yes**. To overcome this edit the character translation table ([option M](#)) for Letter-by-Letter and make sure a comma (character number 45) is translated to ?.

S - Sort "see also" to end as

This option will only have an effect when option **N** above, See also references at end of list is set to **Yes**.

When see also cross-references are needed to be at the end of a list of cross-references, they are normally sorted as *zzz* since it is unlikely that a normal word would have as many as three consecutive Zs in it. However if the Danish or Swedish sort orders are in use, various other characters are sorted after Z, so that the cross references would not come in the right place if subheadings beginning with these characters are present

Danish and Swedish sorting orders added as definition files.

Danish alphabet: a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, æ, ø, å

Swedish alphabet: a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, å, ä, ö

This option allows you to change the way that *see also* is treated by the sort from *zzz* (note that there is a space in from of and after the three zzz) to *ÅÅÅ* (for Danish) and *ÖÖÖ* (for Swedish). Note that these options affect the placement of cross-references within the list of subheadings. To change them within the sequence of page references, see [Merge Options Menu](#), Options **P-R**

T - Hyphen Fix for Chicago-15

In the Chicago-15 sorting convention, a hyphen followed by a number is treated as though it comes ahead of **A** and a hyphen not followed by a number is ignored. If this option is set to **Yes** it looks at the character following a hyphen and if it is a number it replaces the hyphen by @, resulting in the correct sorting order.

The sort Definition file

When the sort program starts up it looks for a sort definition file called DEFAULT8.SRT first in the MACREX [default folder](#) and then in the index folder. If it finds one it loads the definition of the sort automatically. If it does not find one the sort program uses the defaults shown above which are close to British Standard BS3700:1988 (now superseded by ISO999). If you have changed any of the options on the Sort options menu you can save the changes into a file called DEFAULT8.SRT simply by pressing <ESCAPE>. You will then be asked whether you want to save them in the index folder (for the current index only) or the MACREX default folder (for all indexes). If you make changes but do not save them the changes will be effective for the current sort only, allowing you to experiment, because when you leave the sort subprogram the changes will be lost and the defaults will be reverted to on the next occasion. You may also save a sort definition file containing your chosen options using any file name you like by pressing <CTRL>S or F8 from this menu. MACREX will add the extension .SRT to all sort definition files. Any files made this way may subsequently be reloaded by choosing <CTRL>L and entering the file name. This allows you to have several different files available which can then be loaded as needed for the job in hand without having to go to the trouble of redefining everything each time. We also supply sort definition files for various purposes - check your distribution disk for files with the extension .SRT.

Sorting batches

Index entries are stored by MACREX in 10 different files. Each of these contains a batch of entries determined by the letter under which they are filed:

- Batch 1 -A
- Batch 2 -B
- Batch 3 -C
- Batch 4 -D, E and F
- Batch 5 -G, H and I
- Batch 6 -J, K and L
- Batch 7 -M, N and O
- Batch 8 -P, Q and R
- Batch 9 -S
- Batch 10 -T, U, V, W, X, Y and Z
- Batch 11 -queries (not sorted)

In normal use you will not notice this, because the entries are automatically stored in the correct batch when you write to disk or exit from MACREX. Each batch can accommodate a huge number of entries (about 1 million) depending on the capacity of the computer in use, so the total size of the index can be many millions.

Alphabetical sequence used by MACREX

The information given here refers to the *default* word-by-word and letter-by-letter sorts only. Almost any sequence can be achieved by redefining the sort in options [K](#), [L](#), [M](#), [P](#), [Q](#) and [T](#). The entries are not arranged in direct alphabetical order by the sort, but follow the recommendations of BS3700 and BS1749 with certain added features:

- i. Sorting is word-by-word OR letter-by-letter as chosen
- ii. The symbols # ; ' ` [] () \ ~ ^ ! are ignored but the text between them is not.

Text enclosed in tildes (~~) is sorted but not printed.

- iii. Any text enclosed in curly brackets ({}) is either sorted as though it followed the word immediately following the closing curly bracket or ignored entirely by the sort (see [option F](#) above).
- iv. The words a, and, as, for, in, of, on, the, with, (when not at the beginning of an entry, lower case only) are ignored in sorting when the No setting is chosen after the option Sort Prepositions, etc.? (see [option E](#) above). If you would like to change the list of words to be ignored refer to [option K](#) above.
- v. Capital and lower case letters and accented letters are treated equally - if you don't want this you can redefine the translation of the lower case characters (see [option M](#) above, and [Redefining the sort](#) below).
- vi. Certain other symbols (/ - . <space> ^see also^ are given an order of priority which is illustrated below:

```
index
index 69
index, program for 43
index, program for 47
```

the last two entries will normally be merged automatically. In the case of

```
index, tutorial 54
index, ^see also^ catalogue
```

the order will depend on the setting for [Option N](#) - **"See also" references at end of list?**

```
index arrangement 99
index - arrangement 99
```

Note that there is no difference between a dash and a hyphen on the computer keyboard - [see below](#)

- vii. Mc is sorted as Mac, St is sorted as Saint and St. is also sorted as Saint. You may also change these as shown in [above](#).
- viii. A considerable number of users have pointed out that hyphenated words do not appear to sort in the correct order. There seems to be no consensus about the exact positioning of words containing hyphens. For this reason we recommend that you look at the table by calling up [M](#) - Edit character translation table in the Sort Options Menu, read the documentation very carefully, decide whether you want a hyphen equivalent to a space or to be ignored completely and change the character translation table accordingly. An en-dash (ALT-) is sorted as though it were a hyphen. If you wish to ignore a hyphen for a particular entry only you can type

```
Parry{-}Jones
```

If, however, you want the hyphen to be equivalent to a space you can type

```
Parry~ ~{-}Jones
```

Sorting after merging backup files

You can cumulate an index by loading a series of backup files. Providing Autosort is on they will be automatically sorted into alphabetical order after they are loaded. However if you are making a very large index and Autosort is turned off, the entries will not be in alphabetical order and must be sorted. Backup files loaded into a non-autosort index are treated as non-sorted files even if they are already in alphabetical order.

Redefining the sort

The default conventions used in the sort should result in the right order in most circumstances. Use of curly brackets and the sort-force will further allow any sorting order to be achieved. However the sorting section of MACREX has been made completely adaptable so that if you find that you are frequently having to use curly brackets and sort-forces, or if you wish to have different sorting conventions, you can change the sort so that what you want happens automatically.

If you decide to redefine the sort conventions used in MACREX (options K, L and M above) BE CAREFUL! There is no way that the program can check that what you have done will work - it's up to you to get it right. You can always restore the sort to its default state by leaving MACREX and deleting the file DEFAULT8.SRT from the MACREX [default folder](#) and the current [index folder](#).

How the sort works

You only need to read this section if you are going to change the sorting sequence. In order to sort entries the sort program has to have a way of comparing two entries so that it can decide which should precede the other. It does this by comparing corresponding characters in each of the two entries concerned until it finds a difference. It can then compare the two different characters and decide which comes first. Each character (alphabetical, Greek, punctuation mark, symbol or number) is represented in the computer by a number called its ASCII code. In the basic character sets used by MACREX there are 256 characters numbered from 0 to 255. The unicode character set has about 60,000 characters - MACREX can use some of these - see [characters, fonts and accents](#). You can identify the ASCII code for each character on your machine by selecting option [M](#) (Edit character translation table) from the sort menu or by typing <CTRL>U then <ALT>I while editing an entry. The only way that the computer has of knowing which character it is dealing with is by its ASCII code. If the sort simply compared the numerical ASCII code for each character the final order of the index would be strange - all upper case letters would precede all lower case letters so that lower case **a** would follow upper case **Z**, numbers would come at the beginning of the sequence and punctuation marks would be all over the place. To account for this MACREX never compares the ASCII codes themselves. Each character is *translated* to another character by the character translation table before it is used. In this way you can ensure that you get the sequence you want. The table displayed by [option M](#) in the sort subprogram lists 128 characters at a time in eight columns. In each case the ASCII code of the character is shown first, the character which it represents on the PC screen next and finally the character to which it will be translated by the sort for the purpose of comparison. Note that the characters produced by the printer are not guaranteed to be the same as those on the screen. For the ASCII values between 32 and 127 they will probably be almost identical, but the ASCII values above 128 will vary depending on the setup of your computer. It is necessary to use printer and word processor replacement codes to get them to appear on the printer or in the word processor file correctly - these codes are installed by default in MACREX for common setups. There is only room to show 128 of the 256 possible characters on the screen at the same time, but you can toggle between the two halves of the character set by entering **P** at the prompt. If you inspect the table you can see certain things that are very obvious - for example all the upper case letters are translated to lower case and accented letters are treated as their non-accented upper case equivalents. However, there are three features that are less obvious but important to understand if you want to change the table.

- i. Spaces: The space character (ASCII value 32) is not translated by the word-by-word table. This is not because it needs to be sorted as though it had a value of 32 but because it needs special treatment. A space is treated differently according to whether it occurs in the body of an entry or between the end of the entry and the beginning of the page numbers. For this reason you should not change the value of the character to which spaces are translated.
- ii. Ignored characters: Some characters (e.g. quotes) need to be ignored completely by the sort rather than sorted as something else. This is achieved by translating them to an ASCII value of 1. MACREX knows that a 1 should be ignored and if it finds one it will go forward to the next character before making a comparison.
- iii. Zero: The ASCII value 0 (which produces nothing on the screen) has a special meaning to

MACREX. It shows that you have got to the end of the entry. You should never change the translation of 0 - it would almost certainly result in the program crashing. When two entries are compared they are first converted into *surrogate* strings for the comparison. This is done by replacing any of the *text to be replaced* and then by removing any of the ignored words. The two *surrogate strings* are then compared using the character translation table.

Word ignored only at the beginning of a subheading

If you want a word to be ignored only when it is the first word of a subheading or sub-subheading you can achieve this as shown in the following example. Assuming that the word you want ignored is *and*:-

- i. Select L - edit list of changed words and then A - Add replacement to list.
- ii. For the word to be replaced put , and (comma <space>and<space>)
- iii. Make this replaced by , (comma <space>).
- iv. Repeat in the same manner for each word you want ignored in this position.

An example of a complex sorting problem - Icelandic Sorting Order

The following example, while of limited obvious relevance to the average MACREX user, shows how the sort can be manipulated to accommodate something very different from BS3700, and is intended to indicate how such problems might be dealt with in another language.

Iceland uses code-page 861. Icelandic filing order is as follows:

a, á, b, d, eth, e, é, f, g, h, i, í, j, k, l, m, n, o, ó, p, r, s, t, u, ú, v, x, y, ý, þ, æ, ö.

We have provided a sort definition for letter-by-letter sort only (CP861.SRT) which will

- i. Allow several letters to be entered and automatically filed in the appropriate batch. Eth (Ð) is filed with the Ds; thorn, Æ, and Ö are filed with the Zs.
- ii. Sort so that Eth follows D and thorn, Æ, and Ö follow Z in that order.

Notes

- i. There is not at present an automatic way of making words with accented vowels automatically follow those which have no accents. They are normally treated equally. You can make them follow by appending something in tildes at the end of the word, for example ~A~. Thus:

cóir
coir

will be sorted in the order entered, but

cóir~a~
coir

will be sorted correctly.

- ii. You can get header letters for Ð (eth), þ (thorn), Ý, Æ and Ö and blank lines before them by the following method (illustrated for Ð)
 - a. make an entry which reads ~DZZZZ~. This will cause a blank line to be printed at the end of the Ds.
 - b. Make an entry which is simply Ð on its own. This will form a header letter.

- c. If you want a blank line after the header letter make another entry which reads ~DAAAA~.

Sorting defaults

When you receive your copy of MACREX the sort is set up to do the following:

A - First letter sorted **A**
 B - Last letter sorted **Z**
 C - Drive for temporary files **(the drive on which MACREX is installed)**
 D - Merge following sort? **No**
 E - Sort prepositions etc? **No**
 F - Allow for text in curly brackets? **No**
 G - Alphabetization - **word by word**
 H - Sort numbers within entry? **Yes**
 I - Maximum length of sorted numbers **10**
 J - Unconditional re-sort **Yes**
 K - Edit list of ignored words
 L - Edit list of changed words
 M - Edit character translation table
 N - "See also" references at end of list? **Yes**
 O - "See also" text **^see also^**
 P - Sort roman Numbers within entry? **No**
 Q - Replace soft commas? **No**
 R - Soft comma replacement text **>**
 S - Sort "see also" to end as **ZZZ**
 T - Hyphen Fix for Chicago-15 **No**

Macrex Disk Sort

This sort keeps all of the entries on the disk in 10 files (or batches) containing alphabetical groups. Each batch is loaded into RAM and sorted in turn. If one batch is too big to fit into RAM it is split into multiple sections which are sorted separately and then merged together. This allows very large indexes to be sorted. The maximum number will depend on the specifications of the computer being used, but several million should be possible.

While the index is being sorted the messages shown below will appear on the screen. The number of bytes available (displayed on the screen) determines how big a batch has to be before a disk sort is needed. The actual figure may vary according to the amount of random access memory (RAM) in your machine and the number of other memory-resident programs present:

```
Sorting C:TEMP word-by-word
from batch 1 to 14
Press <ESCAPE> or F1 to interrupt sort
300000 bytes available, 723 Entries
Batch 1 already sorted
Batch 2: Number of entries 36
Batch 3: Number of entries 56
32 sorted
```

the above line counts up the number of the batch being sorted If there are more entries in one batch than will fit into the computer's memory at one time, you will see the message

Disk sort operating

The batch will be split up into blocks of temporary files and merged in a final file. The temporary files will be made on the drive you chose on the Sort options menu, or on the same drive as the index if you didn't specify which. The final sorted file has to be made onto the disk holding the index, so there must be spare space on the disk equal to the size of the biggest batch. If you are short of space, delete any files which are not essential, or, if there aren't any, split the index into

sections containing one batch or letter.

If you want to stop the sort press <ESCAPE> or F1 (or <CTRL>Q). You will be asked to confirm that you really want to stop. The batch being sorted at the time you interrupted and all subsequent batches will be left in their original (unsorted) states. The previous batches will be sorted.

Last updated 20 September 2008 [Macrex Help Contents](#)