

TEA&SYMPATHY

Crossword Software for Microsoft Windows

Dictionary Builder Utility

Copyright© 1994-1998 Bryson Limited. All Rights Reserved.

Email: info@bryson.demon.co.uk

Web: <http://www.bryson.demon.co.uk/>

The Dictionary Builder utility converts lists of words into dictionaries that can be used within TEA and Sympathy.

Word lists to be converted must be in a single file, with one entry per line.

To use the builder, follow these steps:

- If no **Alphabet mapping file** is shown, or if you wish to change the one that is shown, click on the **Change...** button and select a TEA&Sympathy Mapping File.
- Click on the **Change...** button for the **Input word list file** and select the word list you wish to convert into a Sympathy Dictionary, then press Ok.
- A default name for the **Output TEA&Sympathy Dictionary file** will be provided; you can override this if you need to.
- You can optionally reduce the range of word lengths included in the dictionary: doing this will speed up the building process and make the output file somewhat smaller.
- You can optionally set the Sort Order.
- Click on **Start**. Dictionary building takes several minutes; the exact length of time taken depends on the size of the word list.
- If you wish to abandon dictionary building, click on **Abandon**.
- A message box will appear when the dictionary is ready. If there are any errors, check the message at the bottom of the main window for the reason. Do not attempt to use partially complete dictionaries with TEA or Sympathy.

Mapping File

TEA and Sympathy need to know how crosswords are constructed from the entries in your word list: the dictionary builder gets this information from a TEA&Sympathy Map File (suffix .TSM) and includes it as part of the TEA&Sympathy Dictionary file (suffix .TSD) that the main programs use.

A number of example mapping files are provided with the release:

default.tsm the dictionary builder automatically loads this file to start with - by default this is the same as **ansi.tsm**

ansi.tsm standard mappings for the Windows ANSI character set - only letters are used to construct crosswords and case and accent differences are insignificant

ansiacce.tsm Windows ANSI character set with distinct accented characters - only letters are used to construct crosswords; case differences are ignored but all the different types of accented characters are considered distinct;

ansicase.tsm Windows ANSI character set with case and accents distinct - only letters are used to construct crosswords, but all the different types of upper and lower case and accented letters are considered to be different and are not allowed to cross with each other;

ansiall.tsm Windows ANSI character set with all characters alphabetic - all characters in the input file are used to construct crosswords, but letter case and accent differences are insignificant;

ansiaca.tsm Windows ANSI character set with all characters distinct - all characters in the input file are used to construct crosswords, and all the different types of upper and lower case and accented letters are considered to be different and are not allowed to cross with each other.

The following table summarizes the combinations available:

<u>TSM File</u>	<u>Accents</u>	<u>Case</u>	<u>Punctuation</u>
ansi.tsm	Insignificant	Insignificant	Ignored
ansiacce.tsm	Significant	Insignificant	Ignored
ansicase.tsm	Significant	Significant	Ignored
ansiall.tsm	Insignificant	Insignificant	Significant
ansiaca.tsm	Significant	Significant	Significant

centeuro.tsm standard mappings for the Windows Central European character set.

baltic.tsm standard mappings for the Windows Baltic character set.

turkish.tsm standard mappings for the Windows Turkish character set.

If you wish to customize your alphabet mappings for a particular puzzle gimmick, or have a word list based on something other than the standard Windows character sets, you will need to create a file in the [TEA&Sympathy Mapping File Format](#) and load it as the mapping file.

Mapping File Format

This section describes the format of the mapping files (suffix .TSM) that the dictionary builder needs to build a TEA&Sympathy Dictionary File.

Mapping files are in a simple format intended to be created and edited within a standard text editor; if the word list being built into a dictionary is in a non-standard character set, then loading a font based on the same character set will make editing the file a lot simpler. The examples below are based on the standard Windows ANSI character set mapping file **ansi.tsm**.

Mapping files consist of two types of lines: either two characters long:

$x=$ indicates that x is a non-alphabetic character

or three characters long:

$x=y$ indicates that x is an alphabetic character equivalent to (and displayed) as y

Take care not to add any characters after the equals (including spaces) in the first type.

If any character 'x' is encountered in the word list that has no entry in the mapping file, then it is assumed to be an alphabetic character equivalent to itself; so leaving out an entry for 'x' is equivalent to having:

$x=x$

Thus **ansi.tsm** includes lines such as:

=
' =
- =

to indicate that space, apostrophe and hyphen are not considered alphabetical and are ignored when a word is entered in a crossword.

and **ansi.tsm** includes lines such as:

E=E
e=E
È=E
É=E
Ê=E
Ë=E
è=E
é=E
ê=E
ë=E

to indicate that all instances of the letter E are considered equivalent in a crossword and displayed as an unaccented capital E.

Sort Order

The dictionary sort order influences the words selected by the automatic grid filler in Sympathy: it provides a coarse adjustment that can be further refined using the **Normal Letter Commonality** setting on the **Filling Properties Dialog**.

In unconstrained situations, the automatic grid filler only needs to read blocks of words from the front of the dictionary files, so the sort order in the dictionary files determines the type of words chosen initially. When the filler has already placed several words to work around, it will read further into the dictionary files, so the sort makes little difference.

The following controls affect the sort order:

Sort. Usually, you will want to leave this checked so that sorting will be done. There are some situations when you may want to uncheck the box so that no sort is done:

- if you wish to create an alphabetically ordered dictionary from alphabetically ordered input (e.g. for use with TEA exclusively);
- if you wish to experiment with your own sort order by constructing a word list in a particular order and getting Sympathy to use that order.

Commonality. Although the default setting of **0** is normally satisfactory, you may wish to adjust this. The value works in the same way as the commonality settings in Sympathy:

- increasing the number will favor words with more common letters: this will reduce fill times, but the fill may include too many "uninteresting" or "crosswordese" words;
- reducing the number will favor words with less common letters; this will increase fill times, but the fill may include more "interesting" words.

The optimum setting depends on the size of the dictionary and the type of grid being filled:

- the smaller the dictionary, the higher the commonality value that's needed to fill a given grid in a reasonable amount of time;
- the more constrained the grid (barred and fully checked grids are highly constrained, whereas blocked grids are relatively unconstrained), the higher the commonality value that's needed to fill a grid in a reasonable amount of time.

Note that duplicate entries in the word list will be eliminated when a sort is done. If sorting is disabled, duplicate entries will be preserved in the dictionary: the automatic grid filler always ensures that there are no duplicates in the grid, but the fill will be faster if there are no duplicates in the dictionary.

