

Term Morphology Editor user manual

Table of contents

1. Introduction.....	1
2. Overview.....	1
3. Installation of Term Morphology Editor	2
3.1. Quick installation using the setup program	2
3.2. Manual installation.....	2
4. Program requirements	3
5. Terms and definitions	3
6. Editing morphology in term bases	4
6.1. Preparation of term base files for Term Morphology Editor.....	4
6.2. Loading of term base file into Term Morphology Editor.....	6
6.3. Editing term morphology	14
6.4. Updating the term base file	22
6.5. Importing the updated term base file	23
7. Application preferences.....	27
8. Program registration	30
9. Additional information	31
9.1. Differences between Morpher.ru and Hunspell morphology engines.....	31
9.2. Specifics in editing term morphology when working with different term base file formats	32

1.Introduction

If you are a user of *memoQ* or *Memsources Cloud*, you may have experienced problems working with terms in Slavic, Turkic and other highly inflected languages or even such languages as Spanish or Italian. Without special preparation of your term bases, *memoQ* and *Memsources* cannot find all inflected forms of your terminology during translation and QA. As a result, translators and editors have to spend additional time to look up terms and can even fail to use proper terminology. To avoid this, you must perform morphological preparation of your term bases in a special way, which is both tedious and time consuming. **Term Morphology Editor** is designed for language service providers (LSPs) and freelance translators who want to save time and effort during preparation of *memoQ* and *Memsources* term bases for better term search during translation and QA.

2.Overview

Terminology databases (term bases) in memoQ and Memsources Cloud provide a significant improvement in translation quality and productivity:

1. The translator can easily insert the translation of the terms found in the source text from the translation results (matches) pane.
2. The translator can insert the term's translation from a dropdown list as they type the translation.

3. Term translation accuracy is checked during the QA stage.
4. memoQ is able to perform automatic repairing of fuzzy TM matches by inserting a translation of a term if the current segment and the TM match are different in a specific term.

In order to take advantage of these terminology-related capabilities, the CAT system must be able to **find any inflected form of a term** in the source segment, and, for QA checks, also in the target segment. To achieve this, especially in highly inflected languages like Russian, the term base must be updated by inserting special wildcard characters – the pipe character (|) in Memsource Cloud and the pipe character (|) and star (*) in memoQ. This is very tedious and time-consuming.

Term Morphology Editor is a program that speeds up editing of term morphology in Memsource Cloud and memoQ term bases. With this program, you can quickly change the way in which the term and its inflected forms are found in text. This applies equally to glossaries which will be imported into a term base for the first time and to existing term bases.

To make your work easier, Term Morphology Editor **provides term morphology suggestions** from two engines – Morpher.ru (a morphological web-service for Russian and Ukrainian languages) and Hunspell (spell checking and morphology system). A terminologist can re-use the morphology suggestions and edit term morphology in a visual editor. As a result, morphological preparation of a term base takes a fraction of the time it would take to edit the term base manually. Once the changes have been made, all you need to do is update the term base file and import it into the term base.

3. Installation of Term Morphology Editor

3.1. Quick installation using the setup program

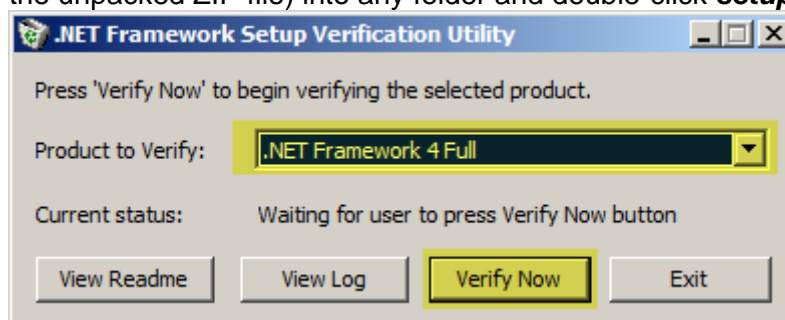
For quick installation of Term Morphology Editor, run “TermMorphologyEditor_Installer_ **version_number**.exe” ([download the latest version](#)) and follow the prompts. You must have administrator rights in order to run the setup program.

To update the program, close it if it is running, and then run the setup program and follow the prompts.

3.2. Manual installation

If you do not have administrator rights or need to install the program for several users, use “TermMorphologyEditor_ManualInstall_ **version_number**.zip” ([download the latest version](#)):

- Unpack the ZIP file into any folder. This folder must be located within a directory to which the user has ‘read’ access rights, for example c:\Users\<user-name>\Documents, or inside a Local Area Network directory with ‘read’ access permissions.
- **The program requires .NET Framework 4.** Most probably, this package is already installed on your computer, however it is better to verify whether it is present before you start using the program. To do this, unpack **netfx_setupverifier.zip** file (located inside the unpacked ZIP file) into any folder and double-click **setupverifier2.exe**.



In the dialogue that opens, choose “.NET Framework 4 Full” from the dropdown list and

click *Verify Now* button. If you see a message “Product Verification Succeeded”, .NET Framework 4 is already installed on your computer. Otherwise you need to download and install it at the following link:

<https://www.microsoft.com/en-US/download/details.aspx?id=17718>

To update the program close it if it is running, and then copy the contents of the updated ZIP file into the program folder, overwriting the existing files.

4. Program requirements

- Microsoft Excel must be installed on the computer.
- Each user must have ‘read’ access rights to the directory where the program is installed.
- The user must have write access to the term base files which will be processed with the program.

5. Terms and definitions

Term base file	A file in a format which can be imported into a new or existing term base.
Term’s morphology pattern	Term text with extra information that specifies which part of each word may change and how the CAT tool must locate the term and its inflected forms in the text
Morphology suggestion, suggested morphology pattern	Term’s morphology pattern proposed by special morphology engines or by the program itself
Fuzzy matching mode (Memsources)	In Memsources – a matching mode which allows the CAT tool to find any inflected forms of the term that are produced by adding letters at the end of any words within the term. For example, “инженер-технолог” – “инженер ^a -технолог ^a ”. This matching mode is used by default if the term is added to the term base without special wildcard characters (see Custom Matching Mode below) and if “exact” parameter is not set to “yes”.
50% Prefix matching mode (memoQ)	In memoQ – a matching mode which allows the CAT tool to find any inflected forms of the term which are produced by adding letters at the end of the last word within the term. For example, “exact match” – “exact match ^{es} ”. This matching mode is used by default if the term is added to the term base without special wildcard characters (see Custom Matching Mode below) and if Exact and Fuzzy matching modes are not applied.
Custom matching mode	A matching mode which allows the CAT tool to find any inflected forms of the term which are produced by changing or adding letters at the end of one or several words within the term. This is typical of highly inflected languages, e.g. when word ending is changed in a different grammatical case or number, e.g. “разн ^{ый} ”. To apply Custom matching mode when editing terms’ morphological settings manually, one must use the pipe character (), and also, for memoQ users, the star symbol (*).

Exact matching mode	A matching mode which specifies that the CAT tool may find only the initial form of the term as specified in the term base. This matching mode is typical of acronyms (e.g., “FAQ”), proper nouns in weakly inflected languages (e.g., “Microsoft Office”), etc.
Fuzzy matching mode (memoQ)	A matching mode used in memoQ which allows it to find additional inflected forms of the term which are produced by adding or changing letters at the end and in the middle of the term with the help of a special algorithm. This matching mode works well with a very limited set of languages, mostly with German.

6. Editing morphology in term bases

6.1. Preparation of term base files for Term Morphology Editor

Term Morphology Editor can process term bases in several formats, each of which has specific requirements:

6.1.1. memoQ

- a) **Microsoft Excel glossaries prepared for first-time import** into memoQ. In this case the term base file must meet the following requirements:
- Processed terms must be in the first worksheet of Microsoft Excel workbook.
 - If a single cell contains several terms, the terms may be separated by any symbol provided that the same symbol is used to separate terms in all other cells and this symbol is not present inside the text of any terms. **Note:** it is not allowed to use the pipe character (|) as the delimiter between terms.

Note that this does not apply to Excel files which were previously exported from memoQ in XLSX format using the appropriate option. Such Excel files cannot be processed with Term Morphology Editor and you should use the CSV format instead as outlined below.

- b) **Term base files in CSV format exported from memoQ** using the Term Base Export command.

Below you can find out how to export an existing memoQ term base for Term Morphology Editor correctly.

First, click *Export Terminology* link in the *Resource Console* or use the same command from the context menu by right-clicking the term base in the *Term Bases* section under *Project Home*.

In the *Term Base Export Settings* dialogue you must use the options outlined in red on the screenshot below:

Term base export settings

File: C:\Users\...\.txt

☒ Export as CSV ☐ Export as XLSX ☐ Export as MultiTerm XML

☐ Include images
Note: images attached to term base entries won't be exported in a plain .csv file.

CSV settings

Encoding: UTF-8

Write field headers in first row: ☒

Delimiter: ☒ Tab ☐ Comma ☐ Semicolon ☐ Other

Export prefix wildcards: ☒

Fields

<input checked="" type="checkbox"/>	Field name	Column header in file
<input checked="" type="checkbox"/>	Entry ID	Entry_ID
<input checked="" type="checkbox"/>	Subject	Entry_Subject
<input checked="" type="checkbox"/>	Domain	Entry_Domain
<input checked="" type="checkbox"/>	Client	Entry_ClientID
<input checked="" type="checkbox"/>	Project	Entry_ProjectID
<input checked="" type="checkbox"/>	Created at	Entry_Created
<input checked="" type="checkbox"/>	Created by	Entry_Creator
<input checked="" type="checkbox"/>	Modified at	Entry_LastModified
<input checked="" type="checkbox"/>	Modified by	Entry_Modifier
<input checked="" type="checkbox"/>	Note	Entry_Note
<input type="checkbox"/>	Image	Entry_Image
<input checked="" type="checkbox"/>	Term definition	[Language]_Def
<input checked="" type="checkbox"/>	Term text (with wildcards)	[Language]
<input checked="" type="checkbox"/>	Term information	Term_Info
<input checked="" type="checkbox"/>	Example of use	Term_Example

Preview

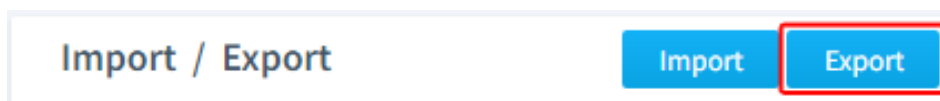
Entry_ID	Entry_Subject	Entry_Domain	Entry_ClientID	Entry_ProjectID	Entry_Created	Entry_Creator
0						
1						

Export Cancel Help

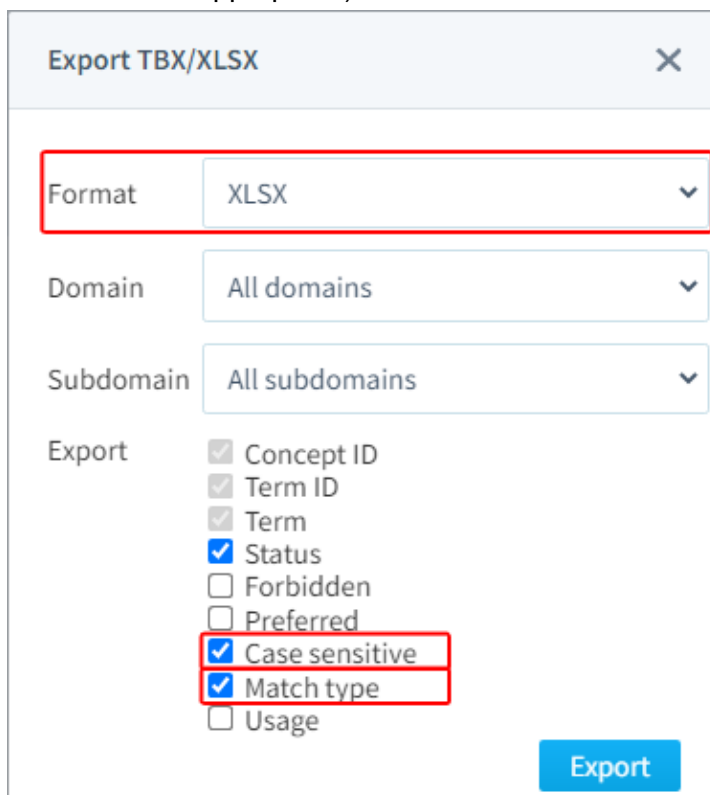
Finally, click *Export* button to save the file to the selected location.

6.1.2. Memsources Cloud

- Microsoft Excel glossaries prepared for first-time import** into Memsources Cloud. In this case the term base file must meet the requirements covered in Memsources Cloud User Manual at the following location: <https://help.memsources.com/hc/en-us/articles/360013696860>
- Term base files exported from Memsources** using the Term Base Export command. Below you can find out how to export an existing Memsources term base for Term Morphology Editor correctly.
First, click the *Export* button on the term base page:



In the pop-up dialogue, select “XLSX” format and make sure that *Case Sensitive* and *Match Type* checkboxes are checked (all the other checkboxes may be checked or unchecked as appropriate):

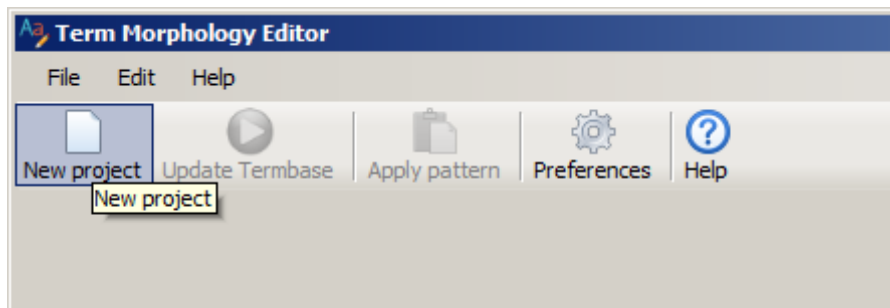


After this click *Export* button and save the downloaded file.

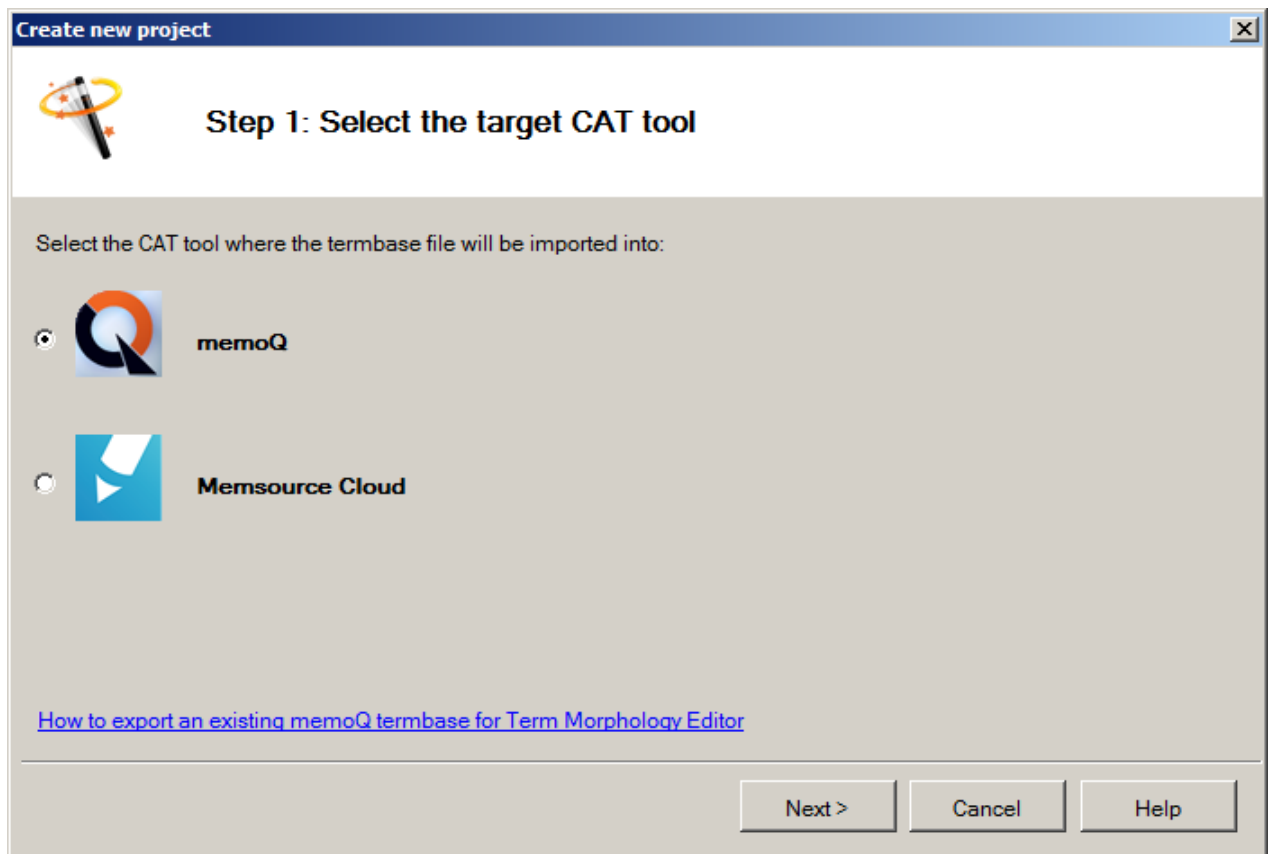
6.2. Loading of term base file into Term Morphology Editor

To open the program run “**Term Morphology Editor.exe**” file (for quick access, you may want to create a shortcut on the desktop or under Windows Start menu).

To start editing a term base file, you must first create a new project. Click *New Project* button or select *File – New Project...* menu:

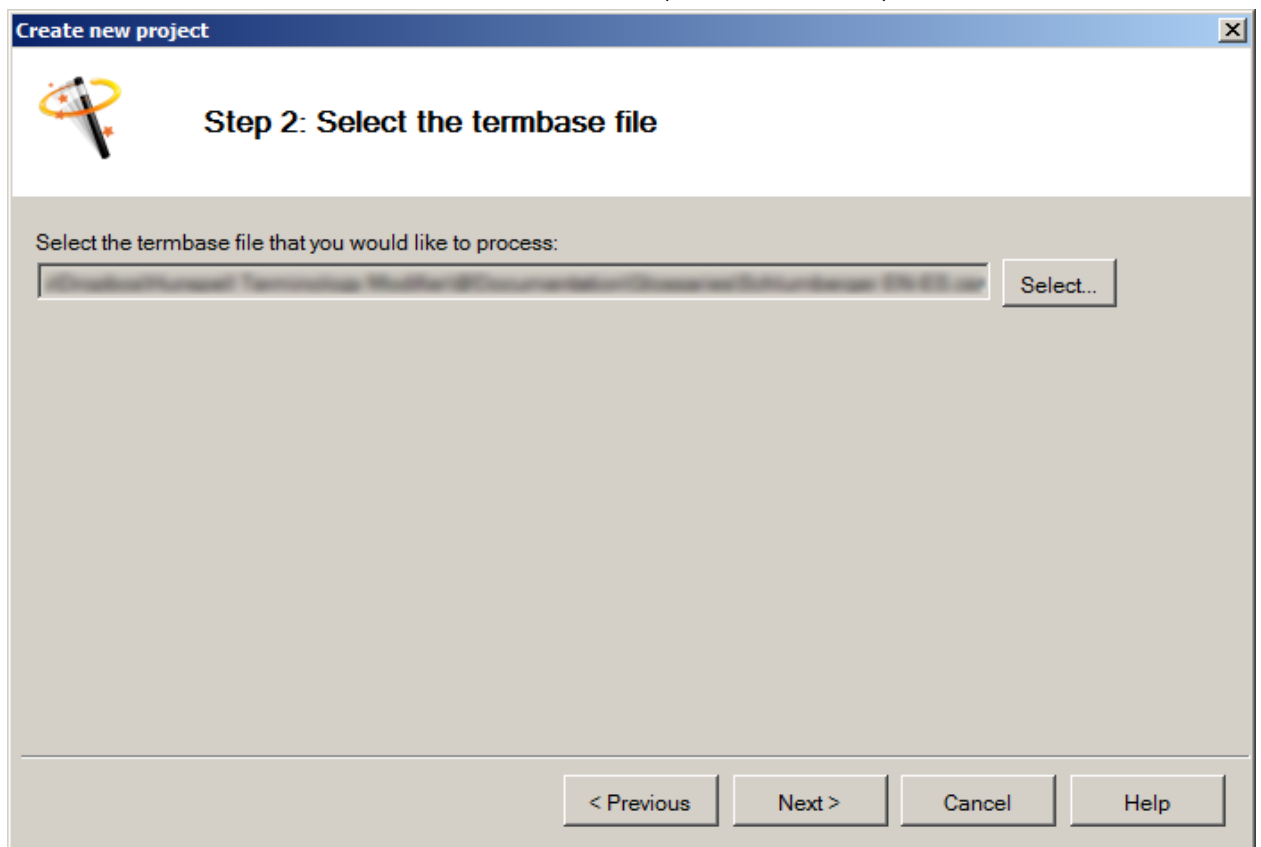


The *New Project Wizard* window will be opened. Select the CAT tool where you intend to import the term base file after processing, and click *Next >* button:



Depending on the selected CAT tool choose the appropriate term base file:

- **memoQ:** file in CSV format (exported from memoQ) or Microsoft Excel format (*.XLS or *.XLSX).
- **Memsources:** file in Microsoft Excel format (*.XLS or *.XLSX).



If you are processing a term base file in Excel format for memoQ you will also need to specify additional parameters:

Select the term base file that you would like to process:

Select...

☒ First row contains column names

Column name	Term language
en	English
es	Spanish
subject	Not a term / Do not process
definition	Not a term / Do not process

Sample text from selected column:

AFE;authority for expenditure
arrival time
acid gas
BOP;blowout preventer
BS&W;basic sediment and water


Check 'First row contains column names' checkbox if the first row contains column names rather than a term entry. In some cases this checkbox will be checked automatically.

For each column that contains terms select the language from the list of supported languages, or select "Not a term / Do not process". To the right of the table you will be able to see sample data from the selected column to help you make the right decision.

When this is done, click *Next >* button.

On the next page of the wizard you will see a list of languages present in the term base file. Check the checkboxes for the languages of the terms you would like to edit, and then click *Next >* button:

Create new project

 **Step 3: Select languages**

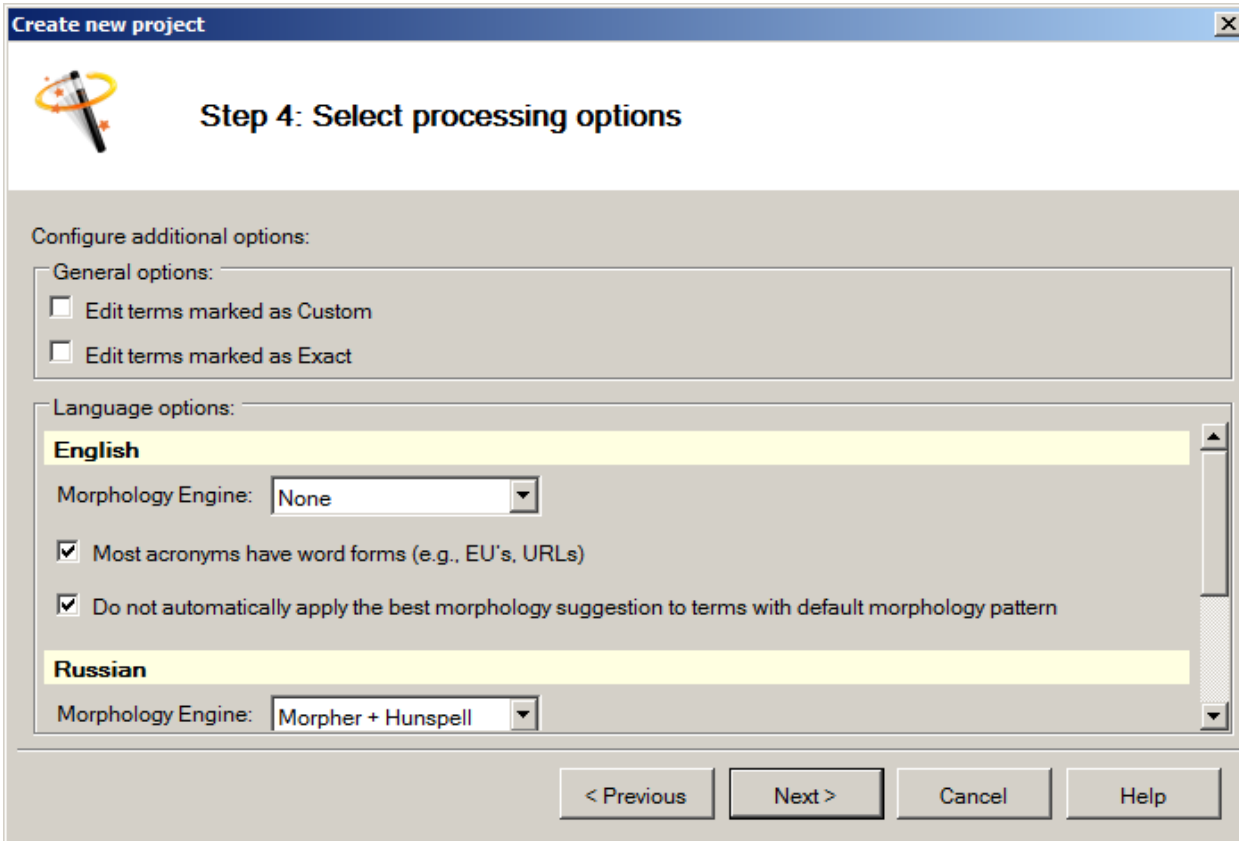
Process terms for the following languages:

☒ English

☒ Russian

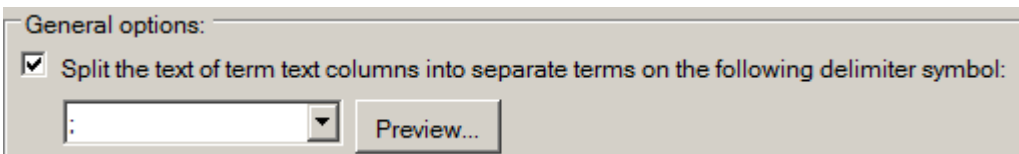
< Previous Next > Cancel Help

On the next page you will need to change additional options for term import and processing:



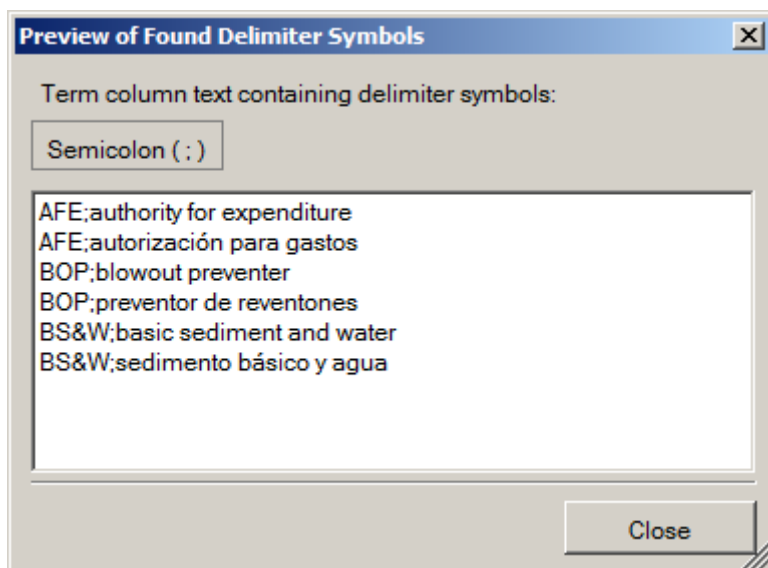
Under General Options, check the checkboxes Edit terms marked as Custom and Edit terms marked as Exact if you wish to edit the morphological settings of the terms which were previously processed by a terminologist, i.e. where the pipe character (|) and star symbol (*) were inserted or the exact matching mode was specified (e.g., for acronyms). When you are processing a memoQ term base file in CSV format you may also check Edit terms marked as Fuzzy checkbox to be able to edit the morphological settings of terms with Fuzzy matching mode.

If you are processing a memoQ term base file in Excel format, the same cell may contain several alternative versions or synonyms of the same term separated by a specific symbol. If you are dealing with such a file, under *General Options* you must check the checkbox “*Split the text of term text columns into separate terms on the following delimiter symbol*” and then select the delimiter symbol from the list:



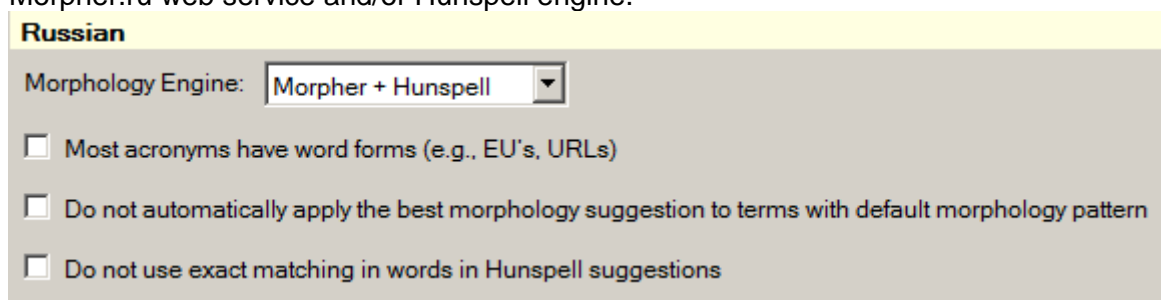
The list will include *Semicolon (;)*, *Comma (,)* and *Slash (/)* if these delimiters are found inside the term text columns. You may also specify a different delimiter symbol. Note that you must not use the pipe character (|) as the delimiter. Also, make sure that the symbol you indicated is only used to separate alternative versions or synonyms and is not found within the text of the terms themselves.

By clicking *Preview* button, you will be able to see which delimiter symbols (out of the three delimiter symbols mentioned above) were found in the term base file:



Under Language options you can change how terms are processed by the morphology engines for each language:

- Select the engines which will be used to provide morphology suggestions from the Morphology Engine dropdown list. For Russian and Ukrainian languages you can select Morpher.ru web service and/or Hunspell engine.



For some other languages, e.g. Spanish, French and Italian, only Hunspell engine is available. If no morphology engine is available for a given language or it is deactivated by default, *None* will be selected in the dropdown list.

- Check the checkbox “Most acronyms have word forms” if the majority of acronyms in this language have inflected forms. For instance, most acronyms in the English language have inflected forms in possessive case (“EU’s strategy”) and in the plural form (“URLs”). If this checkbox is checked, the program will offer a morphology pattern with a variable ending so that the CAT tool can find all the possible inflected forms of the acronym in the text. Conversely, if the checkbox is unchecked and the acronym was not found by the morphology engine, the program will offer a morphology pattern with Exact matching mode, i.e. with only one possible form (initial form).
- Check the checkbox “Do not automatically apply the best morphology suggestion to terms with default morphology pattern (except suggestions from database)” if you are not satisfied with most morphology suggestions from selected morphology engines in the specific language and therefore you would prefer to apply a proposed morphology pattern manually.

If this checkbox is checked, after the term base file is loaded the morphology pattern in *New Morphology Pattern* column for terms with default morphology pattern (i.e., terms without pipe or star characters, without the exact matching mode and without the Fuzzy matching mode in memoQ) will be equal to the initial morphology pattern from the term base.

By default this checkbox is unchecked for most languages. This means that the best morphology suggestion from the morphology engine will be applied to a term automatically if that term has default morphology pattern.

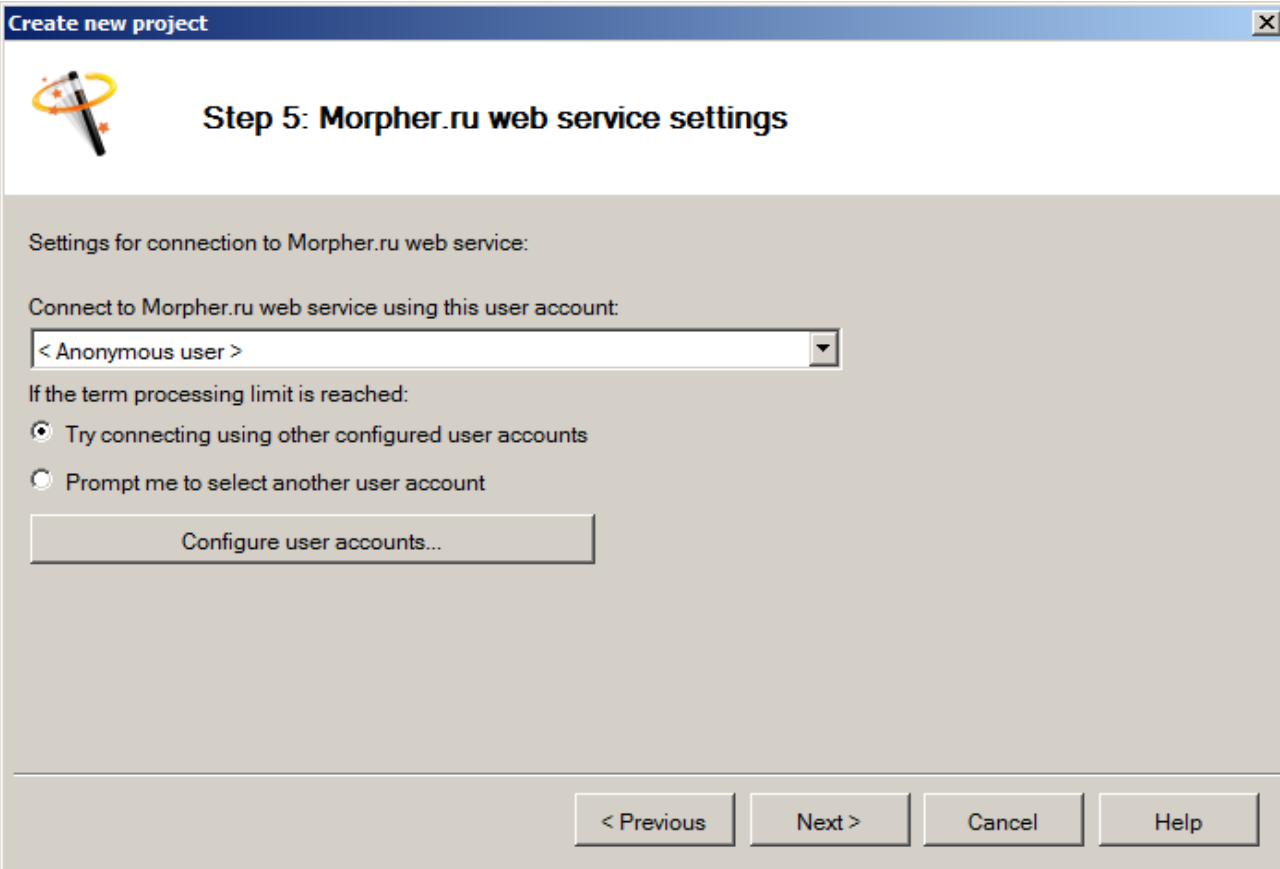
It must be noted that for terms with Custom matching mode, Exact matching mode and memoQ's Fuzzy matching mode the proposed morphology pattern will not be applied automatically regardless of the setting of this option.

- Check the checkbox “Do not use exact matching in words in Hunspell suggestions” if many words or terms are not found in the Hunspell dictionary, causing Hunspell to offer Exact matching mode for a specific term or word. If this checkbox is checked, Hunspell will always assume that such words have word forms which are produced by adding letters at the end of the word. By default this checkbox is checked for most languages with the exception of Russian and Ukrainian.

If you often need to change the language options when creating a new project, you will save some time by changing the default language options using the *Preferences* dialogue.

When you have configured the additional options for term import and processing, click *Next >* button to jump to the next wizard page.

If you selected Morpher.ru morphology engine for Russian or Ukrainian, you will see *Morpher.ru Web Service Settings* page:



Create new project

Step 5: Morpher.ru web service settings

Settings for connection to Morpher.ru web service:

Connect to Morpher.ru web service using this user account:

< Anonymous user >

If the term processing limit is reached:

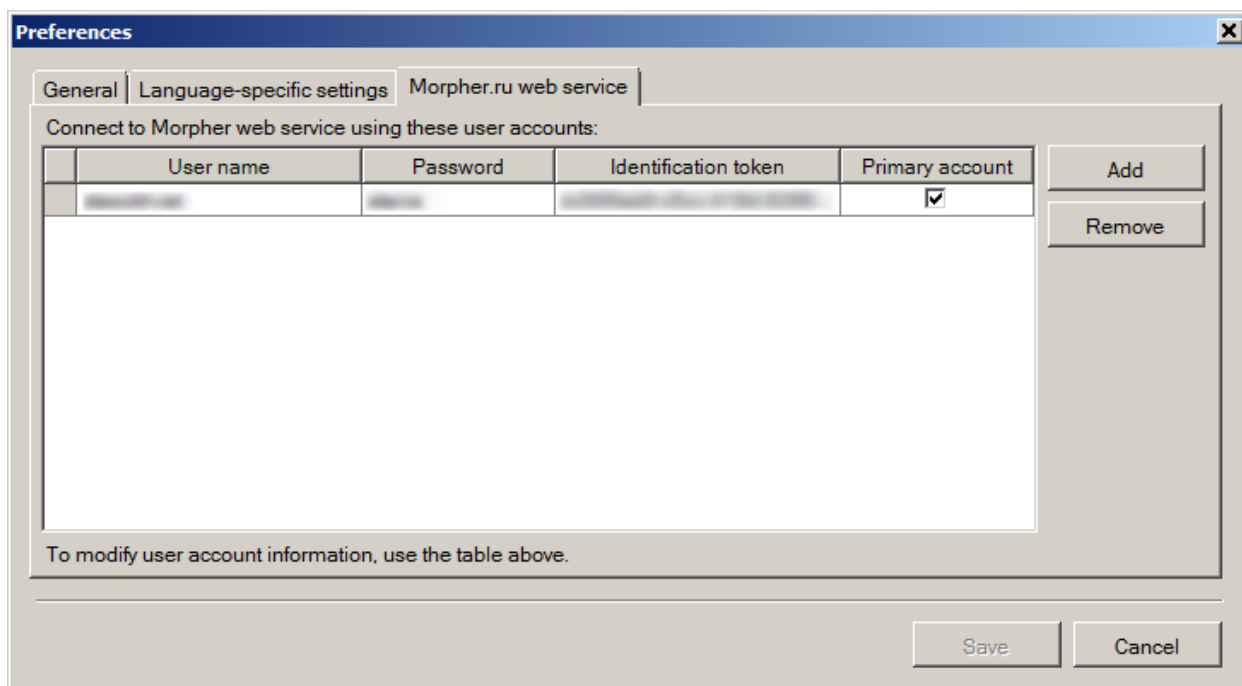
☒ Try connecting using other configured user accounts

☐ Prompt me to select another user account

Configure user accounts...

< Previous Next > Cancel Help

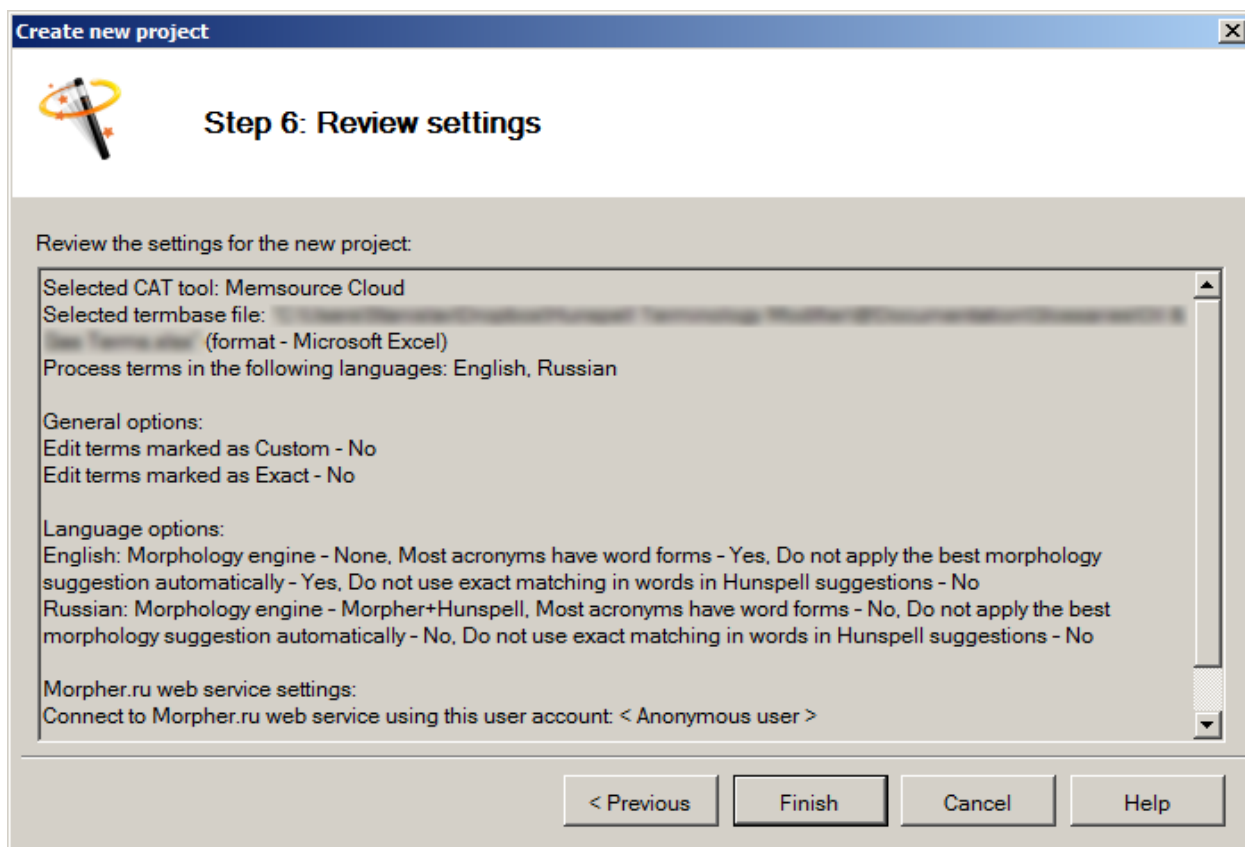
On this page you need to indicate which Morpher.ru user account will be used initially to connect to this web service. Unless you have configured Morpher.ru user accounts the program will connect to the web service under the *Anonymous* user account. Under this user account the program can process up to 100 terms per day for the whole company (which is determined by your IP address). In order to process more terms you must register a user account at <https://morpher.ru/Register.aspx> (the web site is in Russian only) and then specify the user account details in Term Morphology Editor Preferences. Each user account allows to process an extra 100 terms per day free of charge, and if you sign up for paid subscription you will be able to raise this limit significantly (see <https://morpher.ru/ws3/#premium>). To configure user accounts click *Configure user accounts...* button:



Use *Add* and *Remove* buttons to add a new user account to the list or remove it from the list. In order to change user name, password and user identification token you must click a cell and edit its value. In order to obtain the user identification token, go to your user account at Morpher.ru and select the appropriate page, then copy the automatically generated token. It is not necessary to enter the password (to contact the web service, the program uses the identification token), however it may be useful to keep this password here in case you need to access your user account on Morpher.ru. You can also indicate which user account should be the primary one. This user account will be automatically selected in the list on the *Morpher.ru Web Service Settings* page in *New Project Wizard*.

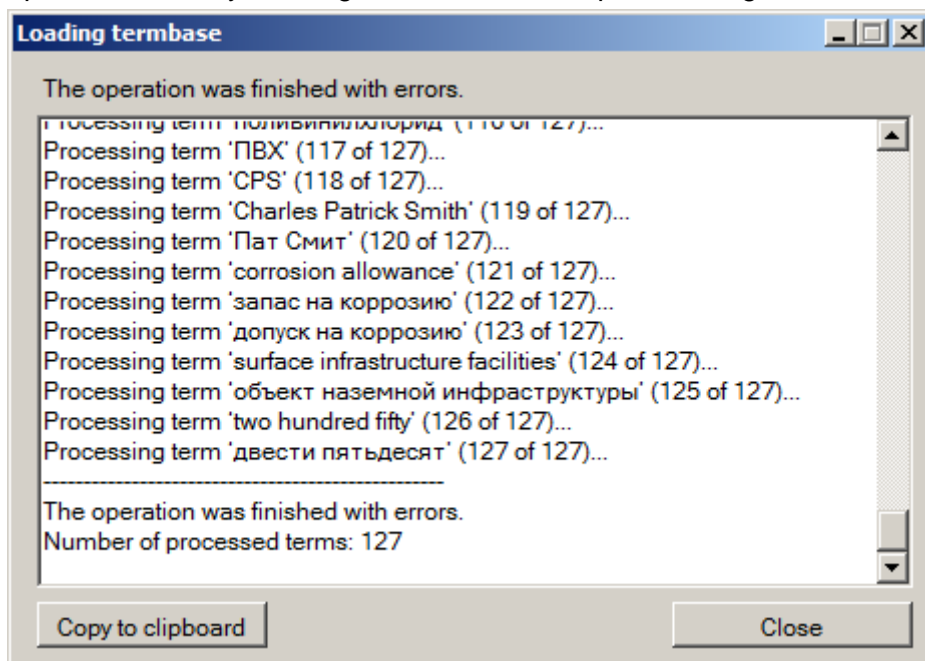
By default, when the term processing limit is reached, the program will switch to the next Morpher.ru user account specified under Preferences or to the Anonymous user account. If you wish to select the next user account manually, check the checkbox "*Prompt me to select another user account*".

Finally, when all the settings have been configured, click *Next >*. You will land on the last wizard page where you can verify the settings of the new project before it is created:



If you need to make any changes, you can go back to previous wizard pages by clicking *< Previous* button. To finish project creation click *Finish* button.

You will see the term base loading dialogue. This dialogue displays the progress of the operation and any warnings / errors that take place during it:



The following errors or warnings may occur during the operation:

- Term '___' was skipped: the term was rejected by Morpher.ru web service.
- Error processing term '___': The term was previously rejected by Morpher.ru web service and was included in the ignore list.
- Error processing term '___': Inflection of numerals is not supported by Morpher.ru web service.

- Error processing term '___': No term is specified.
- Error processing term '___': Morpher.ru daily request limit has been exceeded.
- Error processing term '___': Morpher.ru subscription payment is required.
- Error processing term '___': The user was not found by Morpher.ru web service (incorrect user identification token).
- Error processing term '___': Morpher.ru server error.
- Error processing term '___': Daily limit for identical requests has been exceeded.
- Error processing term '___': Your IP has been blocked by Morpher.ru web service.
- Error processing term '___': Morpher.ru web service does not respond. Connecting again...
- Error processing term '___': There was no response from Morpher.ru web service during 3 previous connection attempts. Morpher.ru web service will not be used to load morphology suggestions during this session.

In the event of errors or warnings displayed in **green** the program will not propose a morphology pattern from Morpher.ru web service and it will propose a suggestion from another morphology engine or use the initial morphology pattern from the term base.

In the event of errors displayed in **red** or **dark-red**, the program will automatically switch to another Morpher.ru user account if this is possible and if the appropriate option was enabled during project creation. When the daily term processing limit is exceeded or if Morpher.ru user identification token is incorrect, you may see a dialogue for selecting another Morpher.ru user account.

When the operation has been completed you can copy the text to the clipboard (e.g., for pasting into Microsoft Word) in order to review the warnings and errors later. When done, click *Close* button.







6.3. Editing term morphology

When the New Project Wizard is closed all the processed terms will be loaded into the table:

The screenshot shows the 'Term Morphology Editor' window. It has a menu bar (File, Edit, Help) and a toolbar with buttons for 'New project', 'Update Termbase', 'Apply pattern', 'Preferences', and 'Help'. Below the toolbar is a table with columns for 'Upd...', 'Pattern', 'Match...', 'Source', 'Old morphology pattern (in termbase)', 'Pattern', 'Match...', and 'Issues'. The table is divided into sections: 'Terms with default morphology pattern' (English), 'Russian', and 'Terms with user-defined morphology pattern (Custom, Exact)'. The 'Russian' section shows terms like 'НО', 'нормально открытый', 'НР', 'нормально разомкнутый', 'регулятор давления', 'предохранительный клапан', and 'полиэтилен терефталат'. The 'Terms with user-defined morphology pattern' section shows terms like 'предохранительный сбросной клапан' and 'бочка с открывающимся верхом'. A dialog box on the right says 'Select a term to view and modify its morphological pattern.' The status bar at the bottom indicates 'Unsaved changes', '5 terms to be updated', and '18 terms (5 terms - suggested pattern, 0 - user pattern, 13 - initial pattern)'.

The table contains the following information:

1	Upd...	New morphology pattern		Old morphology pattern (in termbase)							
2	Pattern	3	Matching type	4	Source	5	Pattern	6	Matching type	7	Issues
Terms with default morphology pattern 8											
English											
	<input checked="" type="checkbox"/> NO		Fuzzy (default)		Tb		NO		Fuzzy (default)		
	<input checked="" type="checkbox"/> normally open		Fuzzy (default)		Tb		normally open		Fuzzy (default)		
	<input checked="" type="checkbox"/> pressure controller		Fuzzy (default)		Tb		pressure controller		Fuzzy (default)		
	<input checked="" type="checkbox"/> safety valve		Fuzzy (default)		Tb		safety valve		Fuzzy (default)		
	<input checked="" type="checkbox"/> pressure safety valve		Fuzzy (default)		Tb		pressure safety valve		Fuzzy (default)		
	<input checked="" type="checkbox"/> pressure relief valve		Fuzzy (default)		Tb		pressure relief valve		Fuzzy (default)		
	<input checked="" type="checkbox"/> open-head drum		Fuzzy (default)		Tb		open-head drum		Fuzzy (default)		
	<input checked="" type="checkbox"/> PETP		Fuzzy (default)		Tb		PETP		Fuzzy (default)		
	<input checked="" type="checkbox"/> polyethylene terephthalate		Fuzzy (default)		Tb		polyethylene terephthalate		Fuzzy (default)		
Russian 9											
*	<input checked="" type="checkbox"/> НО		Exact		M		НО		Fuzzy (default)		
*	<input checked="" type="checkbox"/> нормально открытый		Custom		M		нормально открытый		Fuzzy (default)		
*	<input checked="" type="checkbox"/> НР		Exact		M		НР		Fuzzy (default)		
*	<input checked="" type="checkbox"/> нормально разомкнут		Custom		M		нормально разомкнутый		Fuzzy (default)		
	<input checked="" type="checkbox"/> регулятор давления		Fuzzy (default)		M		регулятор давления		Fuzzy (default)		
*	<input checked="" type="checkbox"/> предохранительный клапан		Custom		M		предохранительный клапан		Fuzzy (default)		
	<input checked="" type="checkbox"/> полиэтилентерефталат		Fuzzy (default)		M		полиэтилентерефталат		Fuzzy (default)		
Terms with user-defined morphology pattern (Custom, Exact) 11											
Russian											
	<input checked="" type="checkbox"/> предохранительный сбросной клапан		Custom		Tb		предохранительный сбросной клапан		Custom		
	<input checked="" type="checkbox"/> бочка с открывающимся верхом		Custom		Tb		бочка с открывающимся верхом		Custom		

- Update column:** If a checkbox in this column is checked, the new morphology pattern will be saved to the term base file if it is different from the old morphology pattern which is already stored in this file. This checkbox is checked by default. Uncheck it if you do not wish to save the changes for specific terms. This checkbox may be checked or unchecked for several terms at once by selecting *Change Update setting for selected term(s)* from the Edit menu.
- New morphology pattern – Pattern column:** this column displays the new morphology pattern of the term which will be saved into the term base file when it is updated.
- New morphology pattern – Matching type column:** this column displays the matching mode which is applied to the term on the basis of the new morphology pattern. For memoQ there are four possible values (see section 9.2): “50% Prefix”, “Custom”, “Exact” and “Fuzzy”. For Memsource there are three possible values (see section 9.2) – “Fuzzy”, “Custom” and “Exact”.
- New morphology pattern – Source column:** this column displays graphically where the new morphology pattern comes from:
 – The morphology pattern was proposed from the database of previously saved terms. The red star * means that the proposed pattern was updated by the user.
 – The morphology pattern was proposed by Morpher.ru web service. The red star * means that the proposed pattern was updated by the user.
 – The morphology pattern was proposed by Hunspell morphology engine. The red star * means that the proposed pattern was updated by the user.
 – The morphology pattern was determined based on the information that most acronyms have or do not have inflected forms in this language based on the language options. This morphology pattern is proposed only if the term contains acronyms and the morphology engines – Morpher.ru and/or Hunspell – are not available for this language.
 – The morphology pattern was taken from the term base file. The red star * means that the initial morphology pattern from the term base was updated by the user. You can see a description of the icon by hovering the mouse cursor over it.
- Old morphology pattern – Pattern column:** this column displays the old morphology pattern of the term which is stored in the term base file.
- Old morphology pattern – Matching type column:** this column displays the matching mode which is applied to the term on the basis of the old morphology pattern.
- Issues column:** for some terms this column may display a special icon  which indicates that the new morphology pattern may have problems. It appears only if the new morphology pattern was proposed by Hunspell morphology engine and may mean two things:

- a) some of the words within the term can be word forms of different words or homonyms which have different morphology model, e.g. the word “*начала*” can have two initial forms – “*начать*” (to start) and “*начало*” (beginning), and each of these two words has different inflections,
- b) the term contains compound words with a hyphen, e.g. “*инженер-технолог*” (process engineer).

If you see this icon for a term, it is strongly recommended to check the proposed morphology pattern.

8. **Terms with default morphology pattern section:** this section displays terms which have a matching mode of “50% Prefix” (for memoQ) or “Fuzzy” (for Memsource), which are used by the CAT tools by default when searching for terms and their inflected forms. In other words, these are all the terms with the exception of terms with pipe or star characters (Custom matching mode), terms with Exact matching mode, and terms with Fuzzy matching mode (memoQ only). Normally, these terms are the ones that require editing during processing of a term base file, therefore they are grouped into a separate section.
9. **Language section:** this section groups all terms in a specific language.
10. **Star symbol in the first column:** the star symbol indicates that the new morphology pattern or the new matching mode of the term are different from the old morphology pattern or matching mode. Therefore, when the Update Term Base command is executed, the term will be updated in the term base file provided that the checkbox in the Update column is checked.
11. **Terms with user-defined morphology pattern (Custom, Exact, memoQ Fuzzy):** this section displays terms whose morphology pattern is different from the default morphology pattern used by memoQ or Memsource, i.e., the term text contains pipe or star characters, the term is marked as Exact (only the initial form is allowed) or as Fuzzy (memoQ only). As a rule, the terms in this section will require less attention. The program proposes morphology engine suggestions for terms in this section, however these suggestions are not applied to the terms automatically.

Terms are displayed in the table in the order they are listed in the term base file. For example, synonyms from the same term entry are located next to each other:

<input checked="" type="checkbox"/>	EMR	Fuzzy...	Tb	EMR	Fuzzy...
<input checked="" type="checkbox"/>	electromagnetic radiation	Fuzzy...	Tb	electromagnetic radiation	Fuzzy...

Morphology engines are not used for the English language, so the table displays the initial morphology pattern from the term base file:

<input checked="" type="checkbox"/>	NO	50%...	Tb
<input checked="" type="checkbox"/>	normally open	50%...	Tb
<input checked="" type="checkbox"/>	pressure controller	50%...	Tb
<input checked="" type="checkbox"/>	safety valve	50%...	Tb
<input checked="" type="checkbox"/>	pressure safety valve	50%...	Tb
<input checked="" type="checkbox"/>	flow meter	50%...	Tb

In order to change a term's morphology pattern you need to select the term in the table. Term morphology editor will be displayed to the right of the table:

Use the following controls to edit term morphology:

1. Use and buttons to change the position where the variable part of the word starts.
2. In order to make a word non-variable (i.e., to allow the initial word form only), keep clicking the button until *No Inflection* checkbox is checked, or check this checkbox manually.

Note the following:


- a. If you are processing a Memsources term base file and the term contains several words, Memsources does not make it possible to make only one word non-variable – the Fuzzy and Custom matching modes in Memsources assume that there may be additional letters at the end of any word.
 - b. If you are processing a memoQ term base file in Excel format you will not be able to check *No Inflection* checkbox for the last word. This is because when such files are imported into memoQ the matching mode of the new terms becomes either *Custom* or *50% Prefix*.
3. In order to make the entire term non-variable, i.e. apply Exact matching mode to it, select “*Exact (variable part not allowed)*” from the dropdown list:

(*Exact* matching mode is not available when processing memoQ term base files in Excel format).

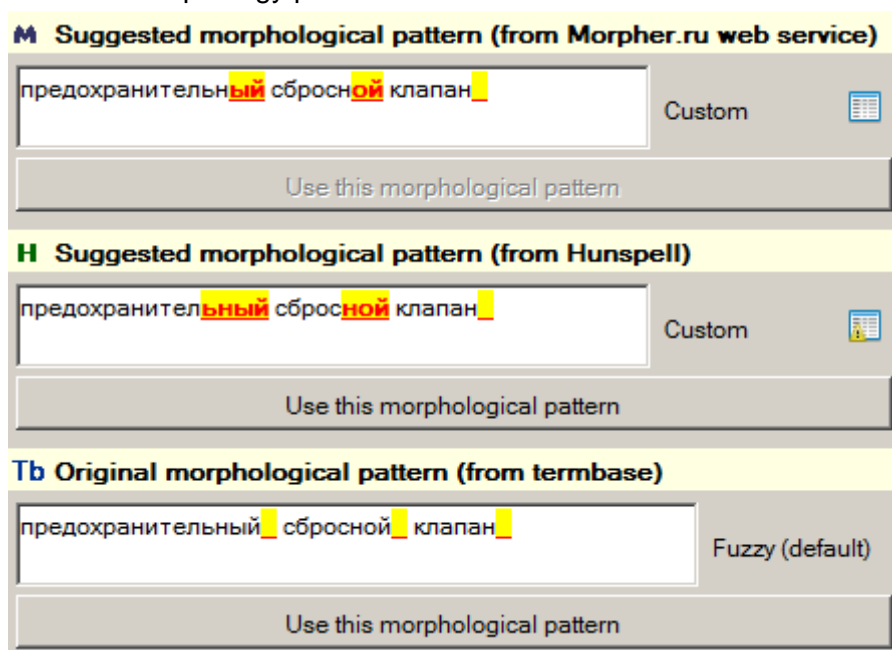
4. To review all possible word forms of a word, hover the mouse cursor over icon to the right of the word:

- The icon is displayed only if the morphology pattern was proposed by Morpher.ru or Hunspell morphology engines.
5. The upper part of the editor displays information about the source of the displayed term morphology pattern.

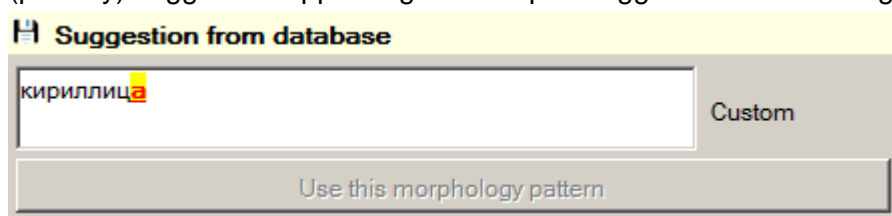
Terms that consist of a single word are displayed somewhat differently:



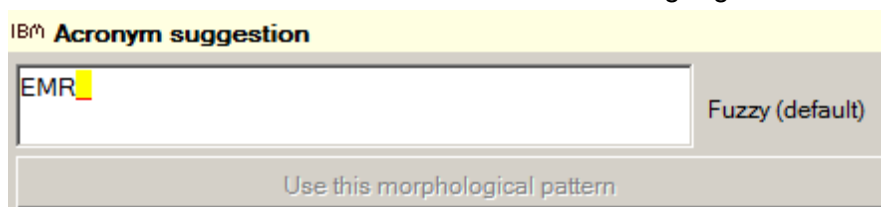
Underneath the term morphology editor you will see suggestions from morphology engines and the initial morphology pattern stored in the term base file:



If the term's morphology pattern was previously saved to the database, it will be the first (primary) suggestion appearing at the top of suggestion list to the right of the term list:



If the term contains acronyms you may see an additional suggestion based on the information whether most terms have inflected forms in the language:

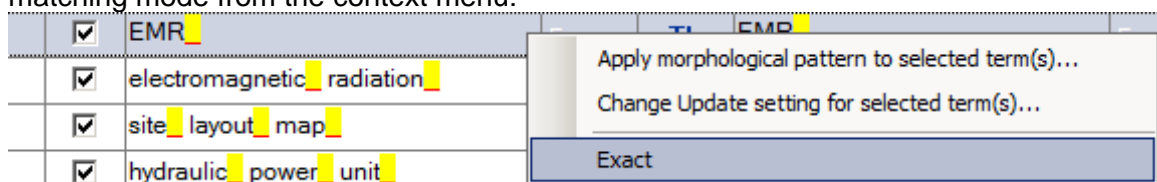


If you would like to apply one of the proposed morphology patterns or make changes to it, click *Use this morphology pattern* button. When you do this, the term's morphology pattern displayed

in the term morphology editor and in *Pattern* and *Matching Type* columns (*New Morphology Pattern* column group) will be updated.

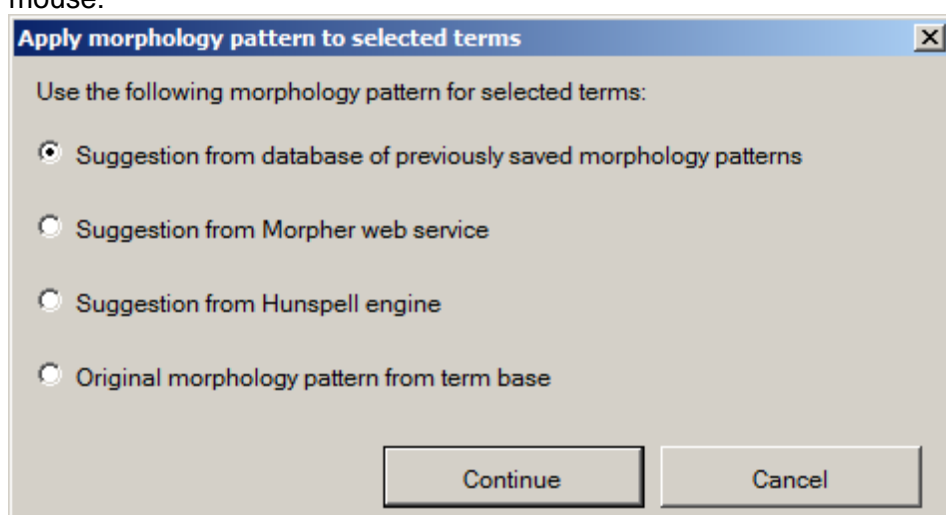
In order to change some parameters of several terms quickly you may use the following commands:

- 1) To change term matching mode, for example apply *Exact* matching mode to several acronyms, select the appropriate terms in the table by clicking them with the left mouse button while holding the Ctrl key pressed, and then right-click with the mouse and select the matching mode from the context menu:

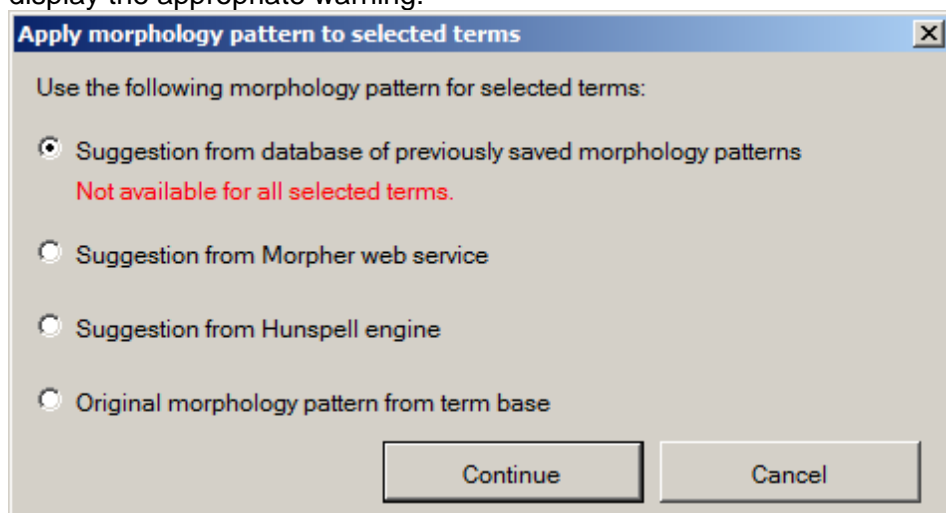


Thus, you can quickly apply *Exact*, *50% Prefix* (default in memoQ), Memsources's *Fuzzy* (default in Memsources) or memoQ's *Fuzzy* matching modes.

- 2) To apply a specific morphology pattern to several terms, e.g. the initial morphology pattern from the term base or a suggestion from the database, Morpher.ru or Hunspell, use *Apply morphology pattern to selected term(s)* command. This command is accessible through the *Edit* menu and via the context menu which opens when you right-click the table with the mouse:

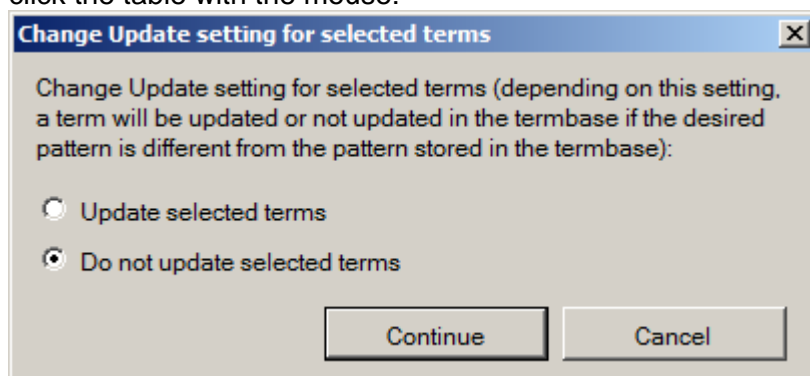


If the selected morphology pattern is not available for all selected terms, the dialogue will display the appropriate warning:



In such cases the program will not update the terms for which the selected morphology pattern is not available.

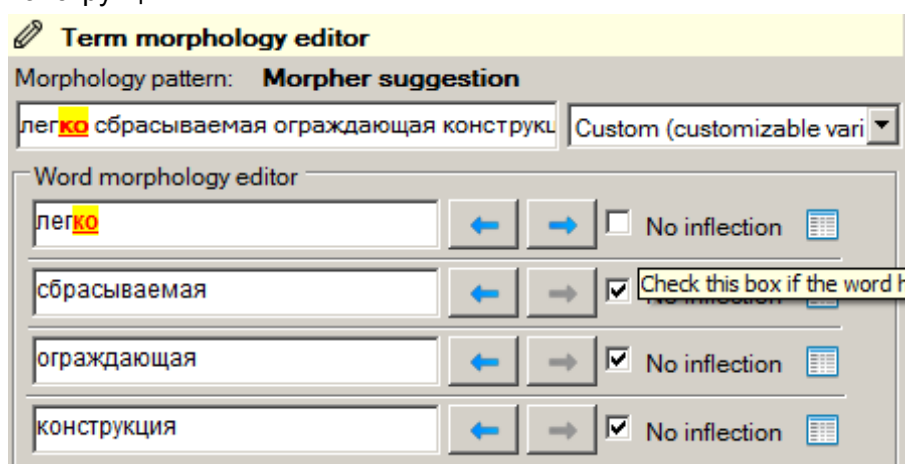
- 3) In order to prevent or allow updating of specific terms in the term base file, use the command *Change Update setting for selected term(s)*. Using this command you will be able to check or uncheck the checkbox in the *Update* column quickly. This command is accessible through the Edit menu and via the context menu which opens when you right-click the table with the mouse:






Example – how to edit an incorrect suggestion from Morpher.ru morphology engine:

1) Editing the morphology pattern of term “легко сбрасываемая ограждающая конструкция” (“blow out panel” in English):

When the project was created, the program automatically applied an incorrect suggestion provided by Morpher.ru web service for term “легко сбрасываемая ограждающая конструкция”:



In order to correct the morphology of this term you need to check *No Inflection* checkbox for the first word or press  button twice for the first word in order to shift the position where the variable part starts. Finally, by using  and  buttons, shift the position where the variable part starts for the remaining words:

Term morphology editor

Morphological pattern: **New (based on Hunspell suggestion)**

легко сбрасываем^{ая} ограждающ^{ая} конструк^и Custom (customizable fixe

Word morphology editor

легко ← → ☒ No inflection

сбрасываем^{ая} ← → ☐ No inflection

ограждающ^{ая} ← → ☐ No inflection

конструкци^я ← → ☐ No inflection

2) Editing the morphology pattern of term “заменить” (to replace):

When the project was created, the program automatically applied an incorrect suggestion provided by Morpher.ru web service for term “заменить” (it must be noted here that Morpher.ru web service cannot handle verbs):

Term morphology editor

Morphological pattern: **Morpher suggestion**

заменит^ь ← → Custom (customizable fixe

M Suggested morphological pattern (from Morpher.ru web service)

заменит^ь Custom

Use this morphological pattern

H Suggested morphological pattern (from Hunspell)


замен^{ИТЬ} Custom

Use this morphological pattern

Tb Original morphological pattern (from termbase)

заменить^ь Fuzzy (default)

Use this morphological pattern

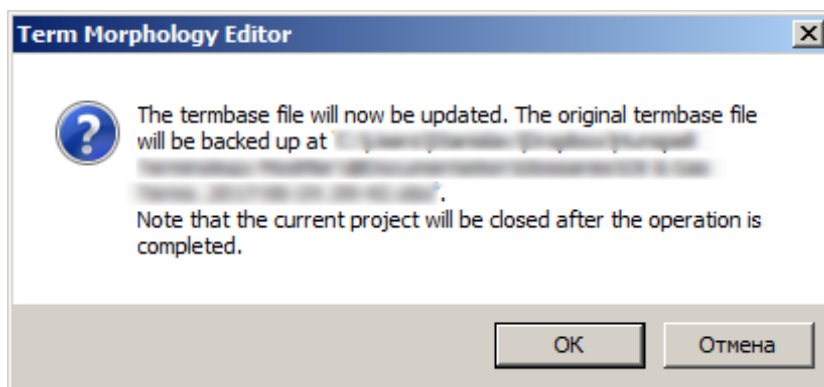
As shown on the figure above, a Hunspell suggestion is also available besides the suggestion from Morpher.ru engine. If you hover the mouse cursor over  icon you will see that all the displayed word forms are consistent with the displayed morphology pattern suggested by Hunspell morphology engine:

H Suggested morphological pattern (from Hunspell)

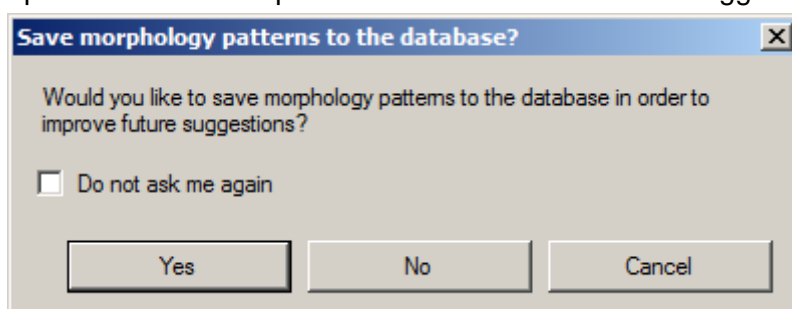
замен^{ИТЬ} Custom

Term inflections:
заменить, замени, замените, заменил, заменила, заменило, заменили, заменю, заменим, заменишь,
заменит, заменят

Therefore, all you need to do is click *Use this morphology pattern* button without any additional editing, or make similar changes in the term morphology editor.



After the dialogue is confirmed, the program will propose to save the morphology patterns of updated terms to a special database which will offer suggestions in the future:



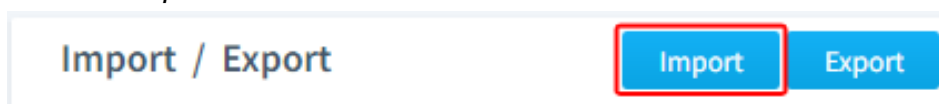
Click *Yes* in order to save the morphology pattern of updated terms to the database, click *No* otherwise (e.g., if you are unsure that the morphology patterns of most terms that you changed are correct), and click *Cancel* to abort the saving process. If you want the program to perform a specific action without displaying this dialogue in the future, check “Do not ask me again” checkbox.

If the term base file contains many terms and you would like to continue editing at a later time, you can do the following: select the terms which have not been checked / edited yet, and then apply the initial (term base) morphology pattern to them by using “Apply morphology pattern to selected terms” command. You can then update the term base file and, when you are ready to continue editing, load this updated version, focusing on terms with default morphology pattern (which are displayed at the beginning of the table).

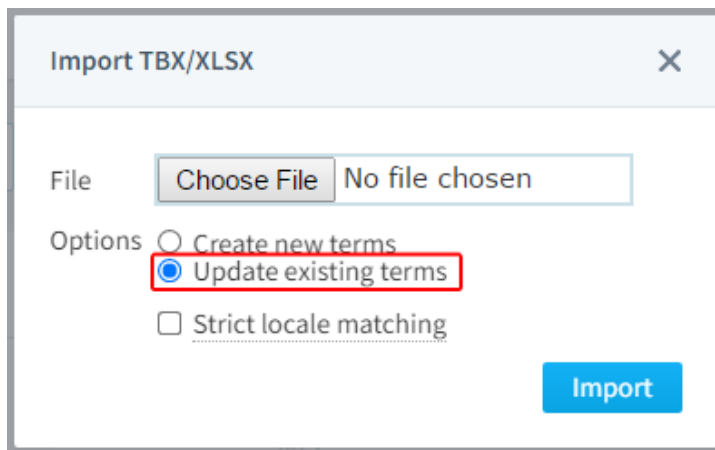
6.5. Importing the updated term base file

6.5.1. Importing into Memsource

After you have updated the term base file, open the term base page inside Memsource Cloud and click *Import* button:



In the dialogue that opens, select the updated term base file using *Select File* button and check *Update Existing Terms* checkbox, and then click the *Import* button.



Attention: unless you check *Update Existing Terms* checkbox, Memsource will add the terms from the file into the current term base, i.e. it will duplicate them, so it is very important to make sure that *Update Existing Terms* checkbox is checked! Otherwise you will have to delete redundant terms manually or create a new term base using the updated term base file.

6.5.2. Importing into memoQ

After updating the term base file, click Import Terminology link in the Resource Console or use the same command from the context menu by right-clicking the term base in the Term Bases section under Project Home, and then select the file.

The process will vary depending on the type of the file you are importing:

- **CSV file**

In the dialogue that opens, check the checkbox “*Update existing entries based on ID*”, and then click *OK* button.

Term base CSV import settings

General

File encoding: UTF-8

Delimiter: ☒ Tab ☐ Split alternatives in field by

☒ Update existing entries based on ID

Fields

Entry_ID, Entry_Subject, Entry_Domain, Entry_ClientID, Entry_ProjectID, Entry_Created, Entry_Creator

Import as term, Import as definition, ☒ Import as other field, Do not import

Preview

Entry_ID	Entry_Subject	Entry_Domain	Entry_ClientID	Entry_ProjectID	Entry_Created
8225	Measurement				10/28/2011 5:
8226	Exploration				10/28/2011 5:
8227	-				10/28/2011 5:

OK Cancel Help

- **Excel file**

In the dialogue that opens, indicate which columns in the file (first worksheet) contain terms and auxiliary information which must be imported into memoQ (see memoQ Help). If some cells contain several terms separated by a specific symbol, check “Split alternatives in field by” checkbox and specify the delimiter symbol. Once all the required settings have been made, click *OK* button.

Term base Excel import settings [X]

General

File encoding: UTF-8

Delimiter: ☒ First row contains field names
☐ Split alternatives in field by: []

☐ Tab
☒ Comma
☐ Semicolon
☐ Other: []

User name where not specified: Stanislav
Subject where not specified: []
Domain name where not specified: []

☐ Update existing entries based on ID

Fields

en
es
subject
definition

☒ Import as term: English
☐ Import as definition: English
☐ Import as other field: []
☐ Do not import

Preview

en	es	subject	definition
AFE;authority for expenditure	AFE;autorización para gastos	[Oil and Gas Business]	A bu...
arrival time	tiempo de arribo	[Geophysics]	The e...
acid gas	gas ácido	[Production Facilities]	A gas...

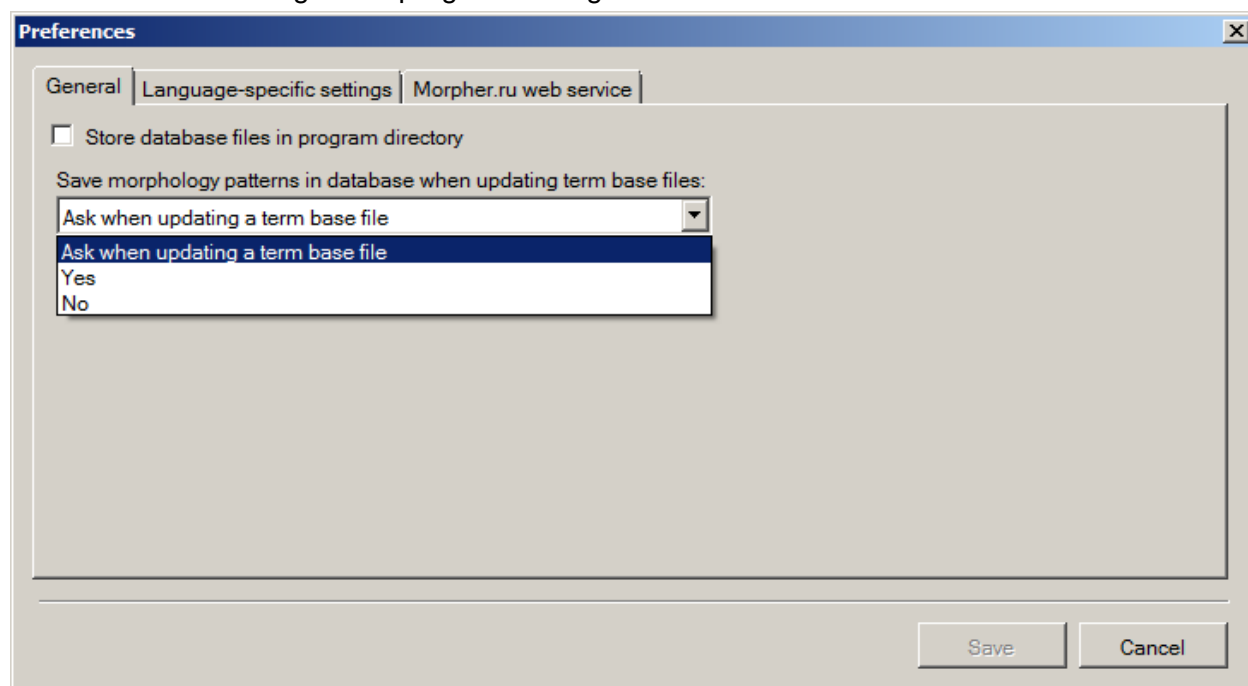
OK Cancel Help

7. Application preferences

To change the program's settings, click the *Preferences* button on the toolbar or use the *Edit – Preferences* menu. The settings are divided into the following sections:

General section:

This section contains general program settings.

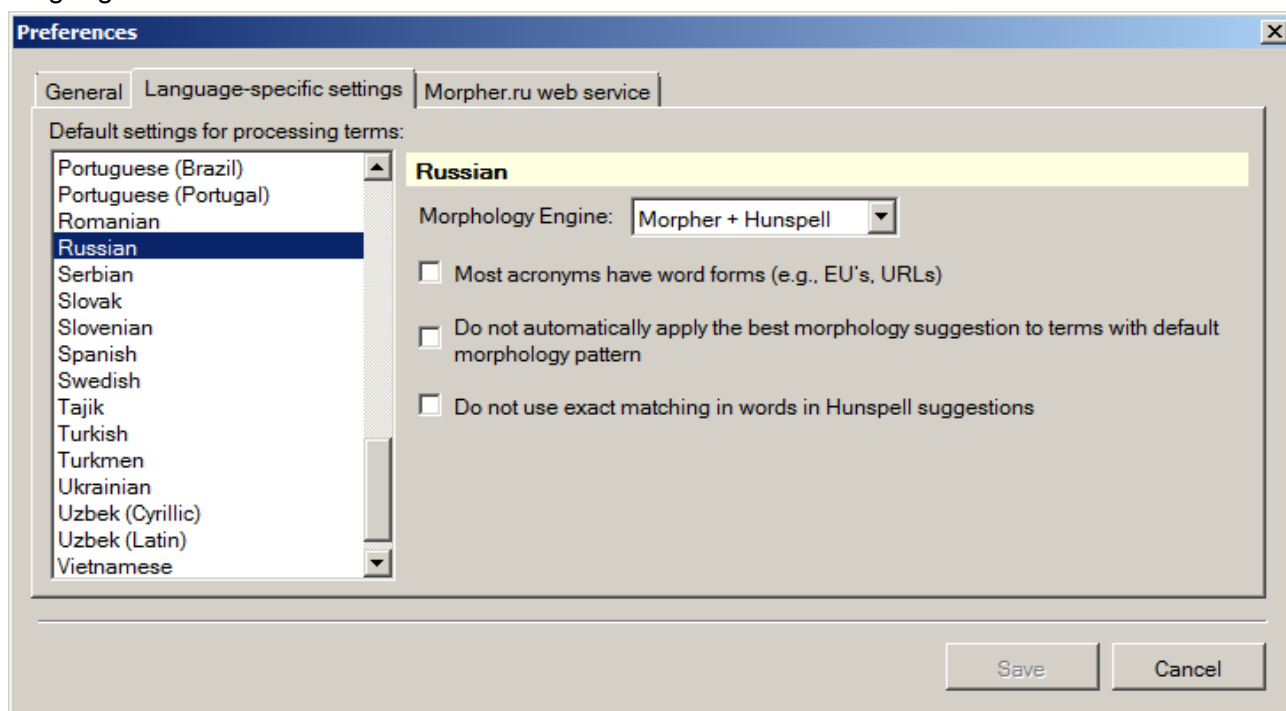


Checkbox “*Store database files in program directory*” specifies which directory must be used to save database files used by the program – in the user directory (“*C:\Users\<user-name>\AppData\Roaming\Term Morphology Editor*”) or in the program’s directory (i.e. in the folder where the program is installed). By default, this checkbox is unchecked, i.e. database files are stored in the user’s personal directory.

Option “*Save morphology patterns in database when updating term base files*” makes it possible to configure how the program should save the morphology patterns of updated terms to a special database when term base files are updated. If you select “Ask when updating a term base file”, the program will confirm saving of morphology patterns every time you update a term base file. If you do not want the program to confirm this operation, choose “Yes” if you want the program to save morphology patterns to the database or choose “No” otherwise.

Language-specific settings section:

In this section you can specify the default settings for morphological processing of terms in each language.



The settings specified here will be used by default when creating a new project (see section 5.2):

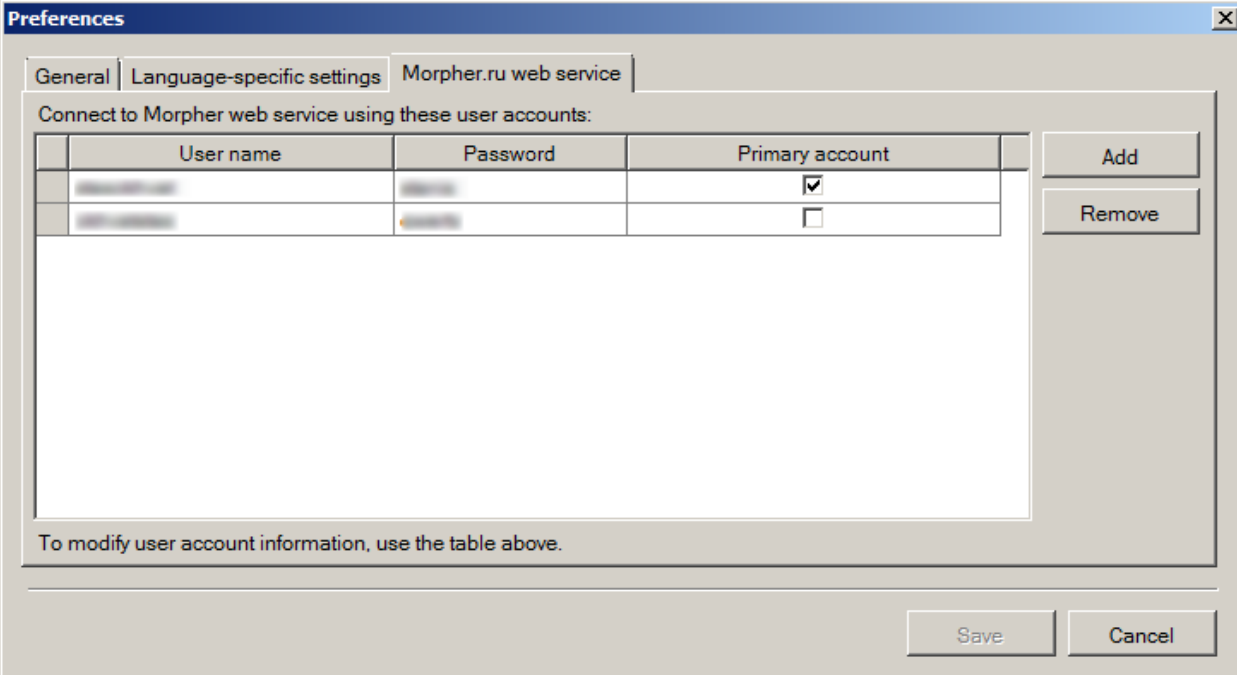
- Select the engines which will be used to provide morphology suggestions from the Morphology Engine dropdown list. For Russian and Ukrainian languages you can select Morpher.ru web service and/or Hunspell engine.
For some other languages, e.g. Spanish, French and Italian, only Hunspell engine is available. If no morphology engine is available for a given language or it is deactivated by default, *None* will be selected in the dropdown list.
- Check the checkbox “Most acronyms have word forms” if the majority of acronyms in this language have inflected forms. For instance, most acronyms in the English language have inflected forms in possessive case (“EU’s strategy”) and in the plural form (“URLs”). If this checkbox is checked, the program will offer a morphology pattern with a variable ending so that the CAT tool can find all the possible inflected forms of the acronym in the text. Conversely, if the checkbox is unchecked and the acronym was not found by the morphology engine, the program will offer a morphology pattern with Exact matching mode, i.e. with only one possible form (initial form).
- Check the checkbox “Do not automatically apply the best morphology suggestion to terms with default morphology pattern (except suggestions from database)” if you are not satisfied with most morphology suggestions from selected morphology engines in the specific language and therefore you would prefer to apply a proposed morphology pattern manually.
If this checkbox is checked, after the term base file is loaded the morphology pattern in *New Morphology Pattern* column for terms with default morphology pattern (i.e., terms without pipe or star characters, without the exact matching mode and without the Fuzzy matching mode in memoQ) will be equal to the initial morphology pattern from the term base.
By default this checkbox is unchecked for most languages. This means that the best morphology suggestion from the morphology engine will be applied to a term automatically if that term has default morphology pattern.
It must be noted that for terms with Custom matching mode, Exact matching mode and

memoQ's Fuzzy matching mode the proposed morphology pattern will not be applied automatically regardless of the setting of this option.

- Check the checkbox “Do not use exact matching in words in Hunspell suggestions” if many words or terms are not found in the Hunspell dictionary, causing Hunspell to offer Exact matching mode for a specific term or word. If this checkbox is checked, Hunspell will always assume that such words have word forms which are produced by adding letters at the end of the word. By default this checkbox is checked for most languages with the exception of Russian and Ukrainian.

Morpher.ru web service section

In this section, you can configure Morpher.ru web service accounts which should be used to access the web service.



Preferences

General | Language-specific settings | **Morpher.ru web service**

Connect to Morpher web service using these user accounts:

User name	Password	Primary account
		<input checked="" type="checkbox"/>
		<input type="checkbox"/>

Add

Remove

To modify user account information, use the table above.

Save Cancel

Use *Add* and *Remove* buttons to add a new user account to the list or remove it from the list. In order to change user name, password and user identification token you must click a cell and edit its value. In order to obtain the user identification token, go to your user account at Morpher.ru and select the appropriate page, then copy the automatically generated token. It is not necessary to enter the password (to contact the web service, the program uses the identification token), however it may be useful to keep this password here in case you need to access your user account on Morpher.ru. You can also indicate which user account should be the primary one. This user account will be automatically selected in the list on the *Morpher.ru Web Service Settings* page in *New Project Wizard*.

If no account is specified in the list, the program will connect to the web service via the anonymous user account.

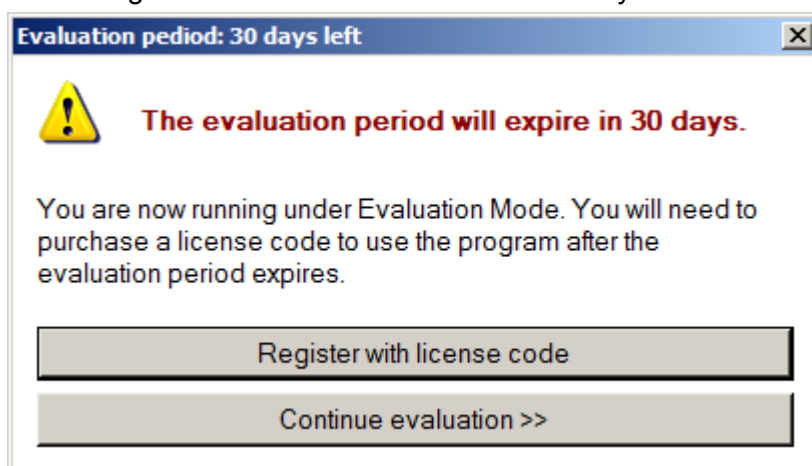
To save the changes you have made to the program's settings, click the *Save* button. To close the window without making any changes to the settings, click the *Cancel* button.

8. Program registration

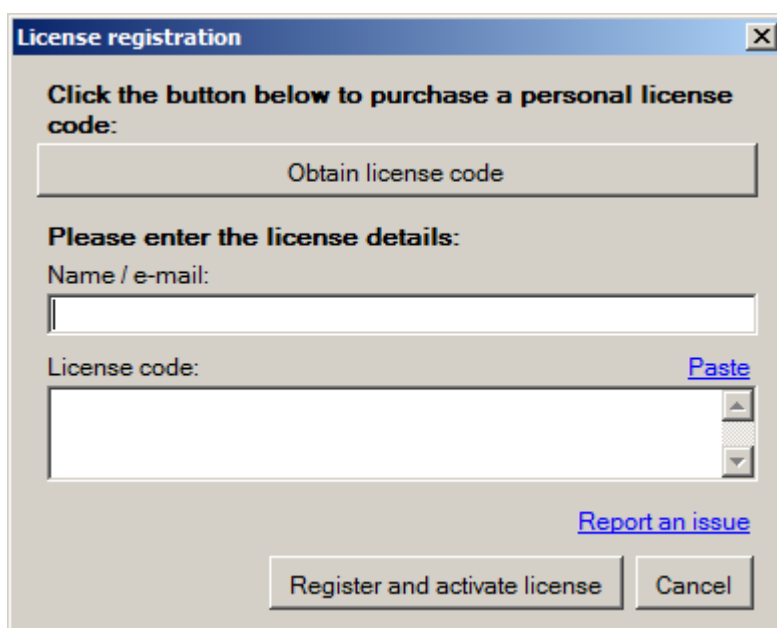
When you open the program for the first time it starts in evaluation mode. During the evaluation period (30 days) you may use the program without any limitations. To use it after the evaluation period expires, you must register the program using a license code. You can purchase a license code through [the website](#) or by contacting us via [this Contact format](#).

When you purchase a license and receive a registration email, register the program in one of two ways:

1. Click “Register with license code” button when you see the evaluation period notification:



The registration dialogue will appear on the screen. Copy the license name and the license code into relevant fields of the dialogue, and then click “Register and activate license” button:



2. Select Help – Register... from the menu and enter the license details in the License Registration dialogue.

According to the license agreement the program may be used on several computers. Individuals may use the program on 3 devices that they own, while corporate users are limited to one work computer and two additional portable computers (laptops). When you enter license details, the program contacts the server and the license is activated on the current device provided that the number of currently activated devices is below the above limit. If you plan to change a work

computer or re-install the operating system, you must deactivate the license on the current device so that you are able to activate it on a new device. To deactivate a license, select Help – Deactivate license... from the menu. The same can be done from About Term Morphology Editor dialogue (select Help – About Term Morphology Editor from the menu).

If it is not possible to deactivate a license on a specific device, contact us via [this Contact form](#).

9. Additional information

9.1. Differences between Morpher.ru and Hunspell morphology engines

There are several differences between Morpher.ru and Hunspell morphology engines:

- Morpher.ru web service:
 - Works with Russian and Ukrainian languages only.
 - Proposes morphological results by considering the inflection of each word inside the combined phrase rather than the inflection of each word separately. For example, in the term “бочка с открывающимся верхом” only the first word has inflected forms, while the other words remain unchanged in any grammatical form of the term, therefore this engine proposes the following morphology pattern: “бочка с открывающимся верхом”. For this reason the morphology pattern proposed by Morpher.ru is predominantly more reliable than the suggestion from Hunspell engine.
 - Cannot inflect numerals written in words, e.g. “одна тысяча двести”.
 - Cannot inflect terms containing verbs, e.g. “наверстать упущенное”.
 - Good at inflecting last names, first names and patronymics as well as rare terms.
 - Works only if there is an Internet connection (however, Internet connection is not required for re-processing the terms which were processed earlier).
- Hunspell engine:
 - Works with the following languages: Albanian, Belarusian, Bosnian (Latin script), Bulgarian, Catalan, Croatian, Czech, English, Estonian, French, Galician, Georgian, Italian, Kazakh, Latvian, Lithuanian, Polish, Portuguese (Brazil), Portuguese (Portugal), Romanian, Russian, Slovak, Slovenian, Spanish, Tajik, Tatar, Turkish, Turkmen, Ukrainian, Welsh.
Much depends on the quality of the Hunspell dictionary. Despite the fact that Hunspell dictionaries are available for a large number of languages, only a limited number of them contain correct morphological information required to make them work with Term Morphology Editor.
 - Offers morphological results of each word inside the term regardless of whether this word has inflected forms as part of the entire term or not.
For example, in the term “бочка с открывающимся верхом” only the first word has inflected forms, however Hunspell proposes the following morphology pattern: “бочка с открывающимся верхом”.
 - Can handle only very common names, and cannot inflect last names, patronymic and rare terms. In other words, in order for Hunspell to suggest a morphology pattern for a word it must be present in the Hunspell dictionary.
 - Participles and verbal adjectives are often treated as though they derived from the corresponding verb, so the non-variable part of the word in a suggested morphology pattern is often identical to the base of the verb, e.g. “соединенный” (deriving from verb “соединять”). This will cause the CAT tool to detect more inflected forms than required (and some of these inflected forms may be unrelated words or phrases).

For Russian and Ukrainian languages it is strongly recommended to select “Morpher + Hunspell” as the morphology engine. This means that the program will receive morphology

suggestions from both morphology engines, and if there is a suggestion from Morpher.ru it will be automatically applied as the *New morphology pattern* for terms with default morphology pattern (i.e., terms whose matching mode is 50% Prefix for memoQ or Fuzzy for Memsource). If you process terms in another language and Hunspell is available for this language, it is recommended to check the reliability of the morphology suggestions offered. If you find out that for many terms, especially those that consist of a single word, Exact matching mode is offered, it means that such words are missing in the Hunspell dictionary or their morphological model is not defined. In such cases you can check the checkbox “Do not use exact matching in words in Hunspell suggestions” when creating a new project. If this checkbox is checked, Hunspell will always assume that such words have word forms which are produced by adding letters at the end of the word. In this way, without any additional editing, the suggested morphology pattern will detect some inflected forms of the term if not all of them.

9.2. Specifics in editing term morphology when working with different term base file formats

memoQ supports 4 term searching approaches (matching modes): *50% Prefix*, *Custom*, *Exact* and *Fuzzy*.

- For terms whose matching mode is 50% Prefix the CAT tool detects the initial form of the term and the inflected forms which are produced by adding letters at the end of the last word (e.g. “дизайн-проект” – “дизайн-проект^а”). At the same time the variable part at the end of the last word in inflected forms must not be longer than the initial form of the term, e.g. “дом” – “дом^{ом}”, but not “дом^{ишко}”.
- For terms whose matching mode is *Custom* the CAT tool considers special wildcard characters “|” and “*” to determine which part of a word can change, and detects inflected forms of the term which meet these criteria. For example, “типичный пример*” – “типичн^{оро} пример^а”.
- For terms whose matching mode is *Exact* the CAT tool detects the initial form of the term only, e.g. “FAQ”.
- For terms whose matching mode is *Fuzzy* the CAT tool finds additional inflected forms of the term which are produced by adding or changing letters at the end and in the middle of the term with the help of a special algorithm. This matching mode works well with a very limited set of languages, mostly with German. This matching mode is displayed in the term morphology editor as follows:



Memsource Cloud supports 3 term searching approaches (matching modes): *Fuzzy*, *Custom* (Fuzzy matching mode with pipe characters) and *Exact*:

- For terms whose matching mode is *Fuzzy* the program detects the initial form of the term and the inflected forms which are produced by adding letters at the end of any words (e.g. “инженер-технолог” – “инженер^а-технолог^а”). At the same time the variable part at the end of any word in inflected forms must not be longer than the initial form of the word, e.g. “дом” – “дом^{ом}”, but not “дом^{ишко}”.
- For terms whose matching mode is *Custom* the program considers special wildcard character “|” to determine which part of a word can change, and detects inflected forms of the term which meet these criteria. For example, “типичный случай” – “типичн^{оро} случа^я”.
- For terms whose matching mode is *Exact* the CAT tool detects the initial form of the term only, e.g. “FAQ”.

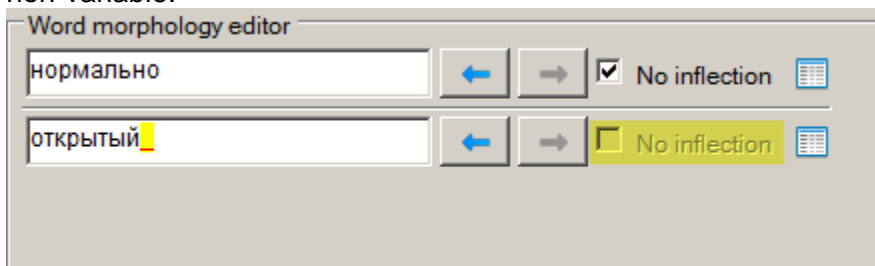
Considering the above, depending on the CAT tool where you are going to import the term base file and also the file format you will notice some differences in morphology editing:

a) memoQ CSV file

When editing term morphology you can select any of the 4 matching modes supported by memoQ and edit the morphology of any word without restrictions.

b) Excel file for memoQ

When editing term morphology it is allowed to use only two term searching modes supported by memoQ – *50% Prefix* and *Custom*. This is related to the way in which memoQ imports files in this format. Therefore, the program will not allow you to make all the words non-variable:



c) Memsources Excel file

When editing term morphology you can select any of the 3 matching modes supported by Memsources. The default matching mode, *Fuzzy*, is displayed as follows:

autorización para gastos

When you edit the morphology of each word the program will not allow you to make a single word non-variable if at least one of the words inside the term is variable.