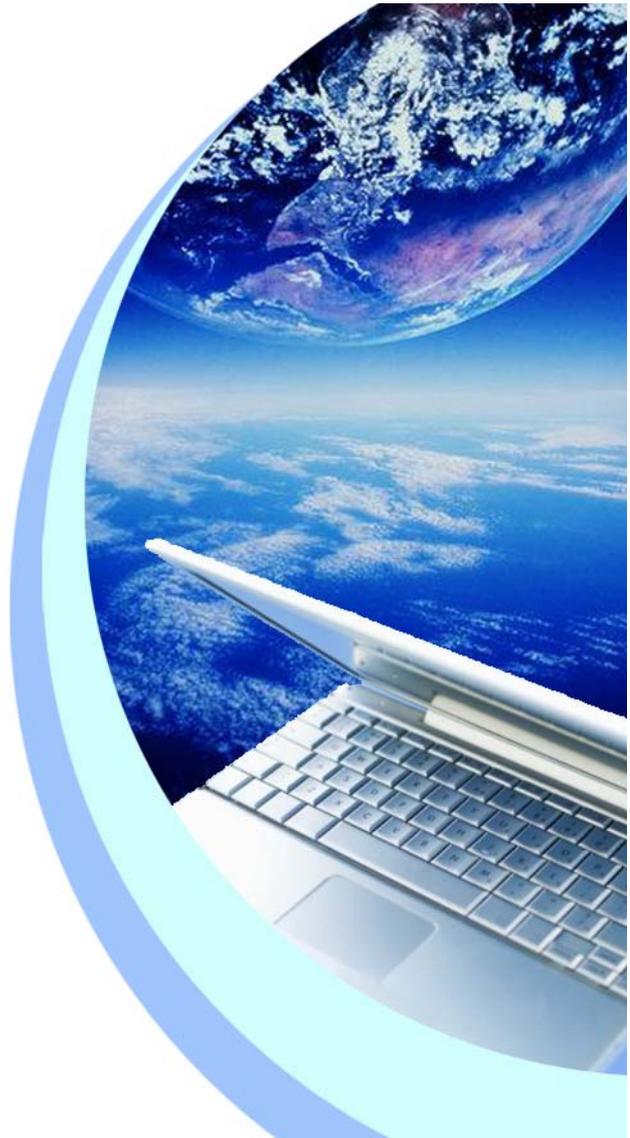




# DashesPro™ for Windows

Version 1.1



## Contents

|  |   |
|--|---|
| Introduction   | 1 |
| Installation   | 1 |
| Quick reference for QuarkXPress  | 1 |
| Quick reference for Microsoft Word   | 1 |
| Hyphenation of files – QuarkXPress tags, InDesign tags, RTF, plain text      | 2 |
| DashesPro™ Menus   | 2 |
| APPENDIX A – About Hyphenation   | 6 |
| APPENDIX B – Stylistics  | 8 |
| APPENDIX C – DashesPro™ Suite: Events and classes for DashesPro™ application | 9 |

## Introduction

The DashesPro™ is designed for hyphenation of texts. It can only be used in conjunction with programs which already support discretionary hyphens. They include Quark XPress (versions 4,5,6) and Microsoft Word (versions 2000, XP and 2003). It can also be used to hyphenate ordinary text files and to unify your hyphenation through applications and platforms.

You can use it to improve hyphenation in your current program and to hyphenate texts in languages not available in your current program.

For the Microsoft Word and Quark Xpress DashesPro™ is available via plug-ins. Both plug-ins are available via menu bar (DashesPro menu item) and have the same menu. By using plug-ins' menu you are able to Hyphenate/Unhyphenate selected text, add words to the exceptions list and select desired language. Comprehensive command set is available only through the main program, which is accessible via tray menu icon (see description of the menu items in the Menu Structure section below).

## Installation

1. Launch 'DashesPro.msi' application and follow the instructions.

## Quick reference for QuarkXPress

1. Launch MS Word.
2. Open the file you want to hyphenate.
3. Select text you want to hyphenate.
4. Select DashesPro | Hyphenate Selection menu.
5. Select the correct language from DashesPro | Languages menu.
6. To remove hyphens from MS Word document select DashesPro | Unhyphenate Selection menu.

## Quick reference for Microsoft Word

DashesPro™ supports hyphenation in the MS Word via clipboard. To hyphenate text:

1. Launch DashesPro™.
2. Open the file you want to hyphenate.
3. Select text you want to hyphenate.
4. Select DashesPro | Hyphenate Selection menu.
5. Select the correct language from DashesPro | Languages menu.
6. To remove hyphens from MS Word document select DashesPro | Unhyphenate Selection menu.

Please note that in order to retain formatting you have to check "Include formatted text in Clipboard" option in the General section of MS Word preferences.

## Hyphenation of files – QuarkXPress tags, InDesign tags, RTF, plain text

To hyphenate a file, select “**Hyphenate File**” under **File** sub-menu of DashesPro™ menu (see menu structure below). Select the file you want to hyphenate. DashesPro™ will insert discretionary hyphens throughout the file. DashesPro™ can only deal with plain text, Rich Text Format (RTF), Quark XPress Tag and InDesign tag files. It will automatically distinguish between RTF and plain text files. If it is an RTF, Quark XPress Tag or InDesign tag file, DashesPro™ will keep the text formatting.]

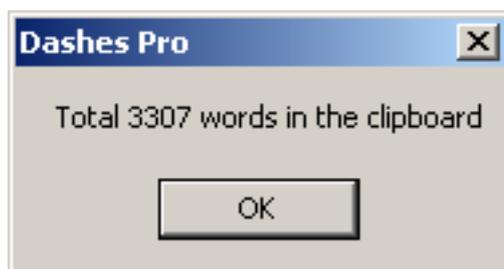
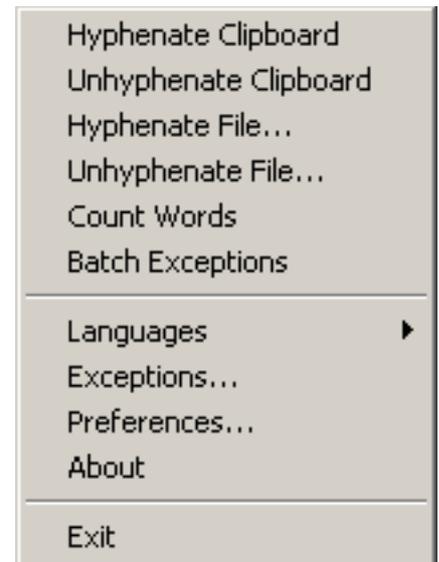
### Menu structure

DashesPro™ menu set consists of DashesPro™ menu, File menu, Edit menu, View menu, Actions menu, Language menu. You can see the description of each point below.

**Hyphenate Clipboard** menu command hyphenates the text on the Clipboard. DashesPro™ tries to preserve as much formatting as possible. It does this by supporting RTF and styled text formats for text it finds on the clipboard. If your application is able to place RTF on the clipboard then all of the formatting is preserved.

**Unhyphenate Clipboard** menu command hyphenates all text in a file. Selecting this menu item opens “Open File” dialog. In the “Open File” dialog locate the file you are looking to hyphenate, choose from the popup menu the correct file type and click Open button. The file will be hyphenated in accordance to the current language selection and other preferences.

**Unhyphenate File...** menu command removes discretionary hyphens from a file. It works the same way as the “Hyphenate File...” menu.

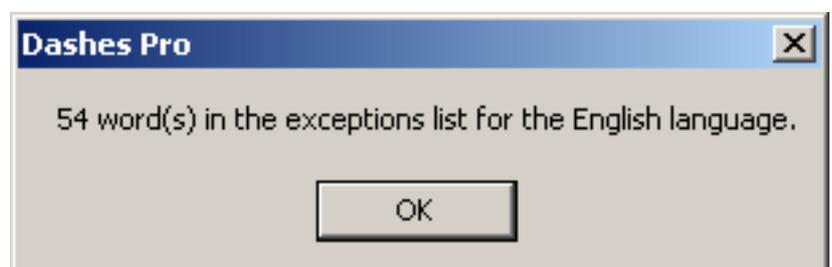


**Count Words** menu command shows the number of words on the Clipboard. Note: a word is a combination of symbols, not separated by white spaces. All the other symbols (such as dots, commas etc.) are counted as the part of the word.

**Batch Exceptions.** If you already have a list of words which you would like hyphenated in a particular way, and you would like to enter these into the exception dictionary all at once do the following. Open the file of hyphenated words. It must have the following format:

William-son  
un-de-ni-abil-ity

In other words, each word you enter must be followed by a carriage return. The hyphens you want entered should appear as hard



hyphens in the text. You may also optionally enter hyphen ranks by inserting the appropriate number after the hyphen such as:

William-1son  
un-1de-2ni-1abil-1ity

Select all the words you wish to have entered into the exception dictionary and copy them to the clipboard. Select “Batch Exceptions” menu item. Your words will be entered into the exception dictionary as you marked them. The “Batch Exceptions” command only enters the word into the exception dictionary if DashesPro™ hyphenates differently from what you have specified in your word list. For large word lists, this cuts down significantly on the size of the resulting exception dictionary and increases the performance without affecting the accuracy.

By default “Batch Exceptions” command does not compare hyphen ranks. That is, if the difference between the word in your hyphenated list and the hyphenation that DashesPro™ assigns is only in the hyphen ranks, then the DashesPro™ will not enter the word into the exception dictionary. For example, DashesPro™ might hyphenate syn-1chro-2nic-1i-3ty and your batch exception list might contain syn-1chro-2nic-2i-2ty. By default DashesPro™ will not add syn-1chro-2nic-2i-2ty into the exceptions list.

Settings menu block consists of **About**, **Preferences**, **Exceptions** and **Languages** items.

## About

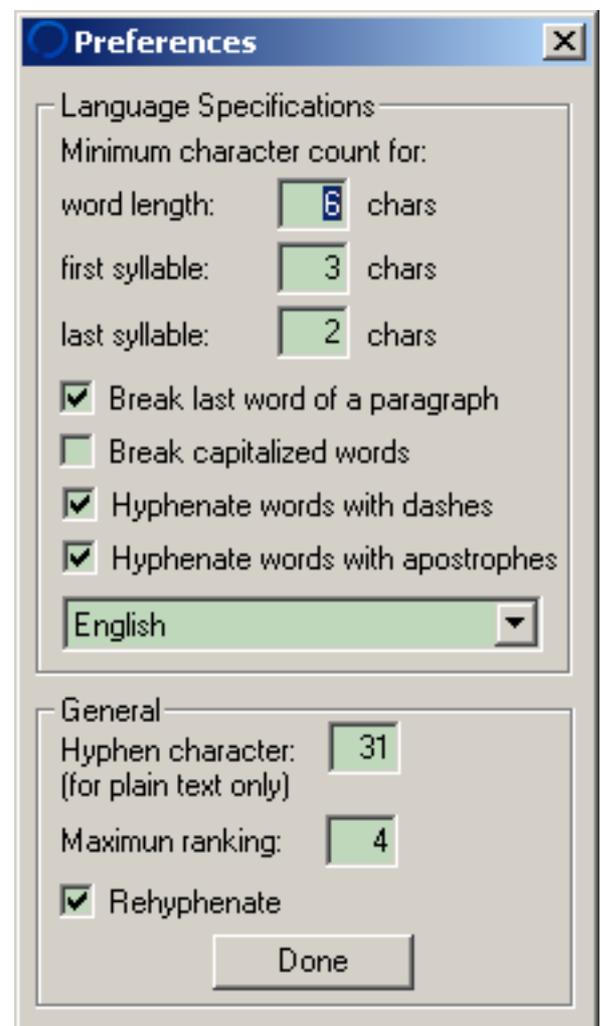
About menu command informs you about DashesPro™ product.

## Preferences

Preferences menu command opens preferences dialog box, which consists of two sections: ‘Language Specifications’ and ‘General’.

In Language Specifications section you can change the following options:

- ‘Minimum character count for:’
- ‘word length’ - sets the minimum number of characters in the words to be hyphenated. The program skips the words with lesser characters.
- ‘first syllable’ - sets the minimum number of main characters for placing the first hyphen in the word;
- ‘last syllable’ - sets the minimum number of ending characters, placing the last hyphen in the word;
- ‘Break last word of a paragraph’ - if selected, the last word of every paragraph



is hyphenated. The end of the paragraph is the unprinted character ‘¶’ (return);

- ‘Break capitalized words’ - if selected, the words beginning with a capital letter are hyphenated;

- ‘Hyphenate words with dashes’ - if selected, the words containing dashes are hyphenated;

Examples: mother-in-law

- ‘Hyphenate words with apostrophes’ - if selected, the words containing apostrophes are hyphenated;

- Current language box – helps you select the language for applying these settings. All changes effect only chosen language.

**Note:** Changing language in this dialog box also changes the current language of the program.

In General settings you can change the following options:

- ‘Hyphen Character’ - chooses a character which is used as the hyphen. This window is used for typing not the character itself, but its decimal ASCII code. This feature works only with plain text. It does not effect the RTF or XPress files or Adobe InDesign tagged text. There are two kinds of hyphens – the ordinary ‘hard’ hyphen, and a discretionary or ‘soft’ hyphen. A discretionary hyphen is shown only at the end of the line. The ASCII code for the soft hyphen varies from application to application. Most applications use code number 31 for the discretionary hyphen. Therefore, 31 is the DashesPro™ default hyphen character code. Usually, 45 is the ASCII code for hard hyphen.

- ‘Maximum ranking’ – DashesPro™ acceptable hyphenation quality. All hyphenations in DashesPro™ are ranked. Rank shows the quality of each particular hyphen. You can vary the maximum ranking from 0 to 4.

- ‘Rehyphenate’ – forces DashesPro™ not to hyphenate those words which it finds already hyphenated.

**Hint:** If some hyphenation seems missing, try to change the maximum ranking.

**Note:** The default rankings are defined (from best to worst) as follows:

0 – word boundary hyphen as in:

white-house

1 – word boundary and prefix or suffix boundary hyphens as in:

under-estimate

relation-ship

2 – hyphen in the middle of a root as in:

estab-lishment

mas-ter

3 – hyphen in the middle of a prefix or suffix as in:

hy-peractive

un-derestimate

4 – hyphen one letter from a word or affix boundary as in:

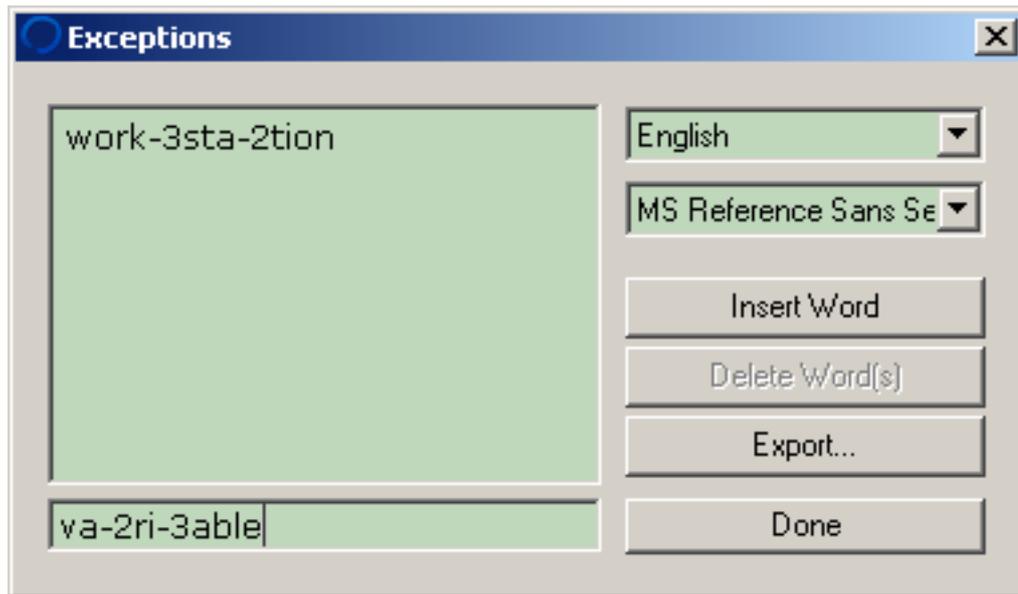
paraphenali-a

o-verreact

## Exceptions

Words in the list are not hyphenated by DashesPro™. Exceptions dictionary is specific for each language. Before making any changes in the Exceptions dictionary you have to make sure the appropriate language is selected. Then you see list of words already inserted in exceptions dictionary of chosen language. You can change font in this window using font selection box.

To add a new word, type it in the line under the list and then click 'Insert Word' button. If this word already exists in the exception dictionary, you are asked if you want overwrite it.



When you add a word to the dictionary, all the necessary hyphenations have to be specified. Before hyphenating you specify the rank of hyphenation with the numbers 0 - 4 (see the Preferences dialog box). Adding hyphenation rank is optional. By default the hyphenation rank is 0.

For example:

in-dus-try

va-2ri-1able

The other way to add a word to the exception dictionary is using 'Batch Exceptions' command (see Batch Exceptions). To remove a word from the dictionary, select the word and click on it and click on 'Delete Word(s)' button. You can select several words at a time by CTRL-clicking on them.

'Export...' button allows you to save your exception list into a new text file. Click on it and after choosing folder, type the file name and click 'Ok' button.

**Note:** Changing language in the exceptions dialog box also changes current language in the DashesPro program.

**Note:** A special warning appears if you try to add to your list a word which already exists in the dictionary.

**Note:** Batch exceptions command checks each word against the way DashesPro™ engine hyphenates this word. So if the word already exists in the exceptions dictionary DashesPro™ will not consider it as an exception.

## Languages

Allows you to select the language of the text which you intend to hyphenate. The current language is marked by '✓' sign. Currently purchased languages are displayed in bold face. Other languages are available in demo mode.



Bulgarian  
Catalan  
Croatian  
Czech  
Danish  
Dutch  
✓ English  
Finnish  
French  
French2

## APPENDIX A – About Hyphenation

People have intuitions about where it makes sense to break a word. The principles behind these intuitions can be defined to a large extent, and dictionaries base their hyphenation suggestions on them. The hyphenations suggested by different dictionaries of the same language vary. The variations depend to some extent on the dialect of the dictionary, but most of the variations are due to ambiguities in our intuitions about where to break a word. We can see this in the fact that all American English dictionaries hyphenate idiosyncratically, as do all British English dictionaries.

### Principles of Word Breaking

There are two principles of word breaking that no dictionary in any language that we know of violates. That is:

B. In a compound word, break the compound between the component words. In other words, breaks like 'an-teater' are universally unacceptable.

C. The unit occurring between two hyphens must be a possible syllable in that language. For example, breaks like:

co-mpatibl-e

whi-tehouse

overdr-essed

are also universally unacceptable.

In many languages, the break does not occur on the actual pronounced syllable boundary. For example, if you were to analyze most people's pronunciation of the word 'ant-eater', we would find that most people pronounce the syllables as 'an/tea/ter'. However, the hyphenation accepted by most dictionaries, namely 'ant-eat-er' still conforms to principle (B), because 'ant', 'eat' and 'er' are all legal pronounceable syllables in English. In the example above, we can see that there are two forces affecting our intuitions about where to break a word. One is that we want to break words on syllable boundaries. The other is that we want to break words between what linguists call morphemes. Morphemes are meaning bearing elements of the word, such as each of the component words of a compound (prefixes, suffixes and roots). A conflict arises when the morpheme boundary is not the same as the syllable boundary. This conflict is solved differently in different languages and by different dictionaries within the same language. This conflict is the main thing that accounts for the discrepancies among and within dictionaries. Let us then add a third principle, which is applied consistently in no language that we know of, but which applies sometimes in most dictionaries for most languages.

#### D. Break words on prefix and suffix boundaries.

Portuguese is an exception to this. In Portuguese, you never hyphenate on prefix or suffix boundaries unless they coincide with syllable boundaries. The English solution to this conflict of syllable and morpheme boundaries is unnecessarily complex, but certain rules have been established by convention, and must be followed for English. The last universal rule is:

#### E. Avoid stylistically poor hyphens like:

##### 1. Hyphens too close to a word boundary, such as:

a-mygdaloid

e-vocative

an-tidisestablishmentarianism

##### 2. Hyphens that fall in the middle of a prefix or suffix:

pol-yethylene

Armeni-an

## Types of Hyphenation Programs

One basic approach to hyphenation is dictionary-based, and the other is algorithm-based. Any algorithm can be made to match any dictionary by adding an exception dictionary to the algorithm which contains all the words that the dictionary and the algorithm hyphenate differently. Thus the real difference between algorithm and dictionary-based hyphenation lies in what they do with words that do not appear in the dictionary. A dictionary-based program does not hyphenate any words that do not appear in the dictionary. The advantage of this is that you are guaranteed no wrong or unreasonable hyphens. The disadvantage is that even in English, a relatively large percentage of the words that actually appear in text are not in the dictionary.

In languages like German and Scandinavian, this is even more crucial, because of word compounding. The compounds in these languages are the longest words, and therefore those you would most like to hyphenate. However since they are freely generated, no dictionary can contain them all.

The two types of algorithms available are what we call statistically based algorithms and linguistically based algorithms. Statistically based algorithms treat hyphenation purely as an information/dictionary compaction problem. The intention is to use various methods to create the most compact data representation that will enable the reproduction of hyphens present in the original list. Linguistically based algorithms are those, which make extensive use of information we have about how words and syllables are structured. Both methods have advantages. However, we do not know of a statistically based algorithm, which is capable of hyphenating Germanic languages with a high degree of accuracy.

It is trivial to write an algorithm that only finds 40%-50% of the possible hyphens accurately. Such algorithms are available at essentially no cost, and they work very fast. However, they tend to put in hyphens at the ends of words, not in the middle where they are most useful. It is easy to imagine how one might write an algorithm that puts in 1% of the possible hyphens with 100% accuracy. You tell the computer when it finds 'ment', 'ments', 'tion', or 'tions' at the end of a word to put a hyphen in front of it. However, it gets exponentially more difficult to maintain a high degree of accuracy as you put in a higher percentage of hyphens. In fact it is impossible for either dictionaries or algorithms to put in 100% of the possible hyphens correctly, because there are a number of words that are spelled the same, but hyphen-

ated differently, such as ‘rec-ord’ and ‘re-cord’. Dashes is a linguistically-based algorithm. It puts in over 99% of the possible hyphens with over 99.9% accuracy in all of the languages offered.

## APPENDIX B – Stylistics

In general, it is recommended that words fewer than 5 characters in length not be hyphenated, and that hyphens not be fewer than 2 letters from the end of the word. In no language are hyphens one letter from the end of a word stylistically acceptable. In German and Finnish, they are simply considered invalid. In other languages, they are of low stylistic quality. These are hyphens with ranking 4 in Dashes, and they should for almost all purposes be removed.

The following rules hold for rewriting characters:

- a. If a hyphen falls between a ‘c’ and a ‘k’ in traditional German, change the ‘c’ to a ‘k’:  
backen → bak-ken
  
- b. In traditional German, as well as in Dutch, Swedish and Norwegian, if three like letters appear in a row due to compounding, one of them is deleted. This consonant is restored when a hyphen is put in.  
Schiffahrt → Schiff-fahrt
  
- c. Dutch has a number of rewriting rules:
  1. Before diminutives ending in ‘-pje’ and ‘-tje’, if the root ends on ‘aa’, ‘oo’, or ‘uu’, a vowel is deleted: autootje → auto-tje, chaltje → chalet-tje, skietje → ski-tje, (cafeetje → caf -tje)
  2. The diaeresis is lost on vowels after hyphenation: zee n → zee-en

In the DashesPro™, the rewriting situations in (a) and (b) are left unhyphenated. In the Dutch examples in (c), the DashesPro™ hyphenates, but does not rewrite.

It is generally not recommended to hyphenate English words with an apostrophe. It is fine to hyphenate French words with an apostrophe. It is generally not recommended in any language to hyphenate words that are already written with a hard hyphen. In all languages it is preferable not to hyphenate proper human names. Other proper names may be hyphenated. Use the Exception Dictionary to prevent hyphenation of proper human names.

We have encountered a number of differences of opinion among typographers and other specialists concerning the hyphenation of French. People differ primarily on whether to hyphenate words such as ‘instruire’ as ‘in-struire’ or ‘ins-truire’. The Canadians in general prefer the latter, but those in France seem to be divided, and feel very strongly one way or the other. For this reason, we have two versions of French, which differ along these lines.

Some examples:

French1 French2

in-1strui-3re ins-2trui-3re

cons-2trui-3re cons-2trui-3re

conta-2mi-2ner con-1ta-2mi-2ner

trans-1ac-2tion tran-2sac-2tion

## APPENDIX C – Dashes Pro™ COM interface

Dashes Pro™ is a COM object which is able to respond on a set of commands. It will allow to use Dashes Pro™ in any automated environments. Interface for Dashes Pro™ is IDashes. It supports the following commands (see sample.vbs the Dashes Pro™ program folder).

**Gets or set Language (language is lowercased)**

```
HRESULT Language([in] BSTR newVal);  
HRESULT Language([out, retval] BSTR* pVal);
```

**Check if the library initialized (language is set)**

```
HRESULT Initialized([out, retval] VARIANT_BOOL* pVal);
```

**Gets the list of supported languages**

```
HRESULT Languages([out, retval] IDispatch** langs);
```

**Hyphenates provided text**

```
HRESULT HyphenateText([in] BSTR textIn, [in] TextFormat fmt, [out, retval] BSTR *pTextOut);
```

**Removes hyphens in the provided text**

```
HRESULT UnhyphenateText([in] BSTR textIn, [in] TextFormat fmt, [out, retval] BSTR *pTextOut);
```

**Hyphenates file**

```
HRESULT HyphenateFile([in] BSTR fileName, [in] TextFormat fmt);
```

**Removes hyphens from the file**

```
HRESULT UnhyphenateFile([in] BSTR fileName, [in] TextFormat fmt);
```

**Hyphenates text in the Clipboard**

```
HRESULT HyphenateClipboard();
```

**Removes hyphens from the Clipboard**

```
HRESULT UnhyphenateClipboard();
```

**Opens Preferences dialog**

```
HRESULT OpenPreferences();
```

**Opens Exceptions dialog**

```
HRESULT OpenExceptions();
```

**Counts words in the text**

```
HRESULT CountText([in] BSTR text, [in] TextFormat fmt, [out, retval] long* pVal);
```

**Counts words in the File**

```
HRESULT CountFile([in] BSTR fileName, [in] TextFormat fmt, [out, retval] long* pVal);
```

**Counts words in the Clipboard**

```
HRESULT CountClipboard([in] TextFormat fmt, [out, retval] long* pVal);
```

**Imports exceptions list from the text**

```
HRESULT ImportText([in] BSTR text, [in] TextFormat fmt, [out,retval] LONG* pVal);
```

**Imports exceptions list from the File**

```
HRESULT ImportFile([in] BSTR fileName, [in] TextFormat fmt, [out,retval] LONG* pVal);
```

Imports exceptions list from the Clipboard

HRESULT ImportClipboard([in] TextFormat fmt, [out,retval] LONG\* pVal);

Helper method to get LanguageID from the language name

HRESULT LanguageID([in] BSTR strName, [out, retval] LONG\* pVal);

Helper method to bring on top the current Dashes window

HRESULT ShowWindow([out, retval] VARIANT\_BOOL\* pVal);