

# **Globalization and National Language Support for Your Version 8.2 SAS Environment**

Historically, SAS has provided options you could use to set up the system to handle data in non-US operating environments. However, these solutions required you to set the options in many different places.

With the 8.2 release, SAS expands and simplifies its support for national languages on all platforms -- from the mainframe to the PC. This applies to data as well as to code, and is especially important to international customers who are running applications in client/server, cross-platform environment. Many features have been consolidated into a simple LOCALE option for a simplified user interface. The FILE, FILENAME and INFILE statements support the new ENCODING= option that enables users to dynamically change the character set encoding for processing external data. Support for Unicode character encoding formats and informats has also been improved. This release also provides production support for the Euro and offers easy conversion for all major European currencies.

8.2 makes internationalization a theme, consolidates language support features, and addresses the problems posed by variant EBCDIC characters. This paper describes the NLS features we have added to version 8.2 and shows some examples of when and how to use those features.

## **Background**

A "locale" reflects the local conventions, language, and culture for a particular geographical region. A locale's conventions may include the formatting of dates, times, and numbers. Locale is not the same as language; a language may be spoken in many countries where conventions are very different. It is also worth noting that a country may have more than one official language. For example, Canada has two languages: English and French.

An encoding is a set of characters that have each been assigned a unique number. The encoding is used by the SAS System to process data. It also influences the TRANTAB system option setting.

In Version 8.1, the locale for the SAS session was controlled by running the Locale Setup Window (LSW) application. The LSW set the TRANTAB system option for the current SAS session, as well as storing locale information in the SAS Registry. That locale information was queried by applications that needed to be aware of the locale name, encoding, Euro character, and translation tables. For example, the SAS session used the trantab information stored in the CORE\LSW\INIT registry was used to set the TRANTAB system option at startup.

Starting in Version 8.2, three new system options are provided to give you more flexibility for setting up the locale and session encoding for your SAS session. The

Locale Setup Window (or LSW) has been redesigned to allow you to enhance the environment set up by the new options.

## **Variant Characters in EBCDIC Encodings**

Among IBM EBCDIC encodings there is a set of 15 problematic characters called 'variants' because they occupy varying code positions:

¢ | ! \$ ¬ | ` # @ ~ [ ] { } \

In other words: the hex values of some special characters in the U.S. English EBCDIC code page have been overlaid or replaced by national characters in other EBCDIC code pages.

Variant characters are used as tokens in SAS syntax, e.g. for character formats, arrays, etc. This caused problems especially when downloading from the mainframe to ASCII-based platforms.

## **The V6 Solution**

The V6 solution was to try to do context sensitive transcoding and to allow user programs to use alternate characters from their keyboards (when for example using a "Å" in place of a "\$" in a SAS program.) Our V6 transcoding often involved proprietary (or non industry standard) translation tables, which force some (but not all) of the variants to transcode in some contexts.

The V6 solutions were not perfect. The two to one mappings of variant characters created headaches throughout the life of V6. For example, the CPORT user needed to remember the source platform in order to determine the proper trantab to use on a CIMPORT.

The V6 solution was expeditious for the V6 product. The most important context sensitive transcodings have been fixed and the others our customers have learned to live with (so what if an "Å" shows up in the log where a "\$" should have been?) Most of the context sensitive transcodings were done in SAS to SAS communications (like in SAS/SHARE or SAS/CONNECT client/server settings or like CPORT files). The fact that SAS was on both ends of the communication allowed SAS programs to know the context of many transcodings.

## **The V8.2 Approach**

In V8.2 we are continuing our implementation of session encoding and addressing the problems posed by compiler variants. The first requirement of a better solution, however, was to run by default in a manner that was compatible with V6. That is, if a user's V6 program reads:

```
INPUT X ÅCHAR10;
```

Then that program must continue to run in V8. The V8.2 approach allows the customer to run in compatibility mode so that existing applications can run unchanged (NLSCOMPATMODE).

The V6 solution was not adequate for V8 and beyond because the V8 SAS system is a more open system than its V6 counterpart. In V8, SAS contends with open clients like

Java, IOM, and SAS/INTRNET. It is no longer reasonable for SAS to know the context of every transcoding or to rely on non-standard translate tables.

A client has the choice of running the system with an explicit session encoding. When they do so, the system will use an internal character representation, which matches the client's native encodings. This means separately encoded MVS versions are shipping in Europe:

```
cp1047 (compiler encoding on MVS)
cp1146 (English U.K.)
cp1147 (French)
cp1144 (Italian)
cp1141 (Austrian/German)
cp1142 (Danish/Norwegian)
cp1148 (Belgian/Swiss-French/Swiss-German)
cp1143 (Finnish/Swedish)
cp1145 (Spanish)
cp870 (Polish et. al.)
```

SAS needed a solution for variant characters because the V6 solutions were not adequate for all of the V6 problems. In addition, the number of problems with the V6 context sensitive approach will grow as SAS gets more into the open interface arena. The open market expects software to get and return data in standard encodings. The SAS system must use standard encodings to effectively exchange data with non-SAS software. Ambiguity of code points (like "Å" and "\$") must be removed in the long term. The 8.2 approach is a first step to getting there. We expect most of our V8 customers will run in compatibility mode. However, over time we expect the norm to switch to run in a standard session encoding that matches the client's OS encoding.

The system options NLSCOMPATMODE and NONLSCOMPATMODE will help to make a transition as smooth as possible.

## **NLS-related System options**

Three new NLS-related system options were added to Version 8.2. The LOCALE option allows you to set the locale for your Version 8.2 SAS session and sets the ENCODING, DFLANG, and TRANTAB system options. The ENCODING option sets the encoding that SAS uses for processing external data. This option also sets the TRANTAB system option.

NLSCOMPATMODE allows you to run existing SAS programs, providing compatibility with previous versions of the SAS system. The NLSCOMPATMODE option is set by default and most users will not need to change the option setting.

All of these options are valid in the configuration file and at session startup. They are documented in the SAS System Help for the host environment where you run the SAS System. More information is provided below.

### **NLSCOMPATMODE**

The NLSCOMPATMODE system option is used to determine the encoding used by SAS for processing data. There are two settings:

**NLSCOMPATMODE: providing compatibility with previous releases of the SAS system**

When the option is set to NLSCOMPATMODE, you can run SAS production programs that you created prior to version 8.2 without needing to make any changes. In "NLS compatibility mode", SAS uses the compiler encoding, or the encoding used to compile the SAS system on your platform, for processing external data and SAS syntax.

This compiler encoding is also the same encoding used for US English and the encoding used if you do not change the locale or encoding of the SAS session. The compiler encoding was also used by SAS in Version 6 and Version 8 for all data processing.

NLSCOMPATMODE is the default setting for the option.

**NONLSCOMPATMODE: providing support for non-English data processing**

NONLSCOMPATMODE turns off NLS compatibility mode. For NONLSCOMPATMODE, all character data will be processed using the ENCODING option encoding, including external data, SAS syntax, and user data. In this mode, the encoding that SAS uses to process character data is the encoding set by the ENCODING or LOCALE option. If neither of these options has been set, SAS will continue to use the compiler encoding.

For information on how the ENCODING option interacts with NLSCOMPATMODE and NONLSCOMPATMODE, see the section "How to configure your system" below.

**LOCALE**

The LOCALE option allows you to specify a locale setting for your SAS session. When LOCALE is set, the DFLANG, TRANTAB and ENCODING options are also updated to match the locale you selected.

LOCALE option values contain the language name. For some locales, you can also specify a country name or region to provide more specific locale information. For example, valid locales for France are French and French\_France. See the SAS System Help for the host environment where you run the SAS System for values that can be specified for LOCALE.

When LOCALE is set, the DFLANG system option is set to a value that corresponds to the chosen locale or "English" if no corresponding value is available. For more information about the DFLANG option, see the *SAS Language Reference*.

The locale you set also has a common encoding that is used most often on the platform where the SAS System is running. When the LOCALE option is set, the ENCODING option is set to match this common locale encoding and also sets the TRANTAB option to support the locale encoding. However, if the ENCODING system option is also specified at startup, the encoding in the ENCODING option will take precedence over the locale encoding and will become the new SAS encoding for the session.

Unlike the LSW, the LOCALE option does not store values in the SAS registry. Run PROC OPTIONS to display the value of the LOCALE option.

## ENCODING

The ENCODING option sets the encoding that SAS uses for processing external data. The ENCODING option can be set by specifying ENCODING at startup or, more commonly, by specifying a LOCALE that uses an encoding other than the default compiler encoding. Valid values for ENCODING are listed in the SAS System Help for the host environment where you run the SAS System.

The NLS compatibility setting does not change the ENCODING option value. However, if you are running SAS in NLS compatibility mode, the ENCODING option encoding is used by a few services. For example, HTML and JavaScript output generated by ODS will be in the ENCODING option encoding. SAS Integration Technologies also uses this encoding when connections are established with client applications.

**Note:** ENCODING values on one platform are not necessarily supported on another platform.

When ENCODING is set, the TRANTAB option is also set. On most platforms, all of the first five trantabs are set:

- local to transport
- transport to local
- upper case
- lower case
- character classification trantabs

For more information about the TRANTAB option, please refer to the *SAS Language Reference*.

### *Special trantab settings on EBCDIC platforms*

On EBCDIC platforms, such as the OS/390, additional trantabs are set in the TRANTAB option. In NLS compatibility mode, a special scanner trantab is set as the 6th trantab in the TRANTAB option. This trantab reassigns variant characters from value in the ENCODING option encoding to value in the compiler encoding. This reassignment only occurs for characters in your SAS syntax, not for characters that are in quoted strings or data lines.

This special scanner trantab allows you to use the characters from your keyboard for SAS syntax instead of making the character substitutions you were required to make in Version 6 and Version 8. However, the character substitutions are supported, so your existing SAS applications will still run. For example, if a Danish user running SAS on OS/390 selects a locale of Danish and NLSCOMPATMODE, that user can use either the '\$' or the 'Å' where SAS syntax calls for a '\$'.

If you choose to run SAS with NLS compatibility mode turned off, all of the characters in your SAS syntax must be in the encoding of the ENCODING option. No trantab is set in the 6th slot of the TRANTAB option. However, a scanner classification (SCC) trantab is set when compatibility mode is off, which is set as the eighth trantab in the TRANTAB option. This SCC trantab tells the scanner the correct classification, or type, of each character it finds in your SAS syntax. For example, the scanner assigns a special classification to the '\$' character. That classification needs to be assigned value that matches the '\$' character in the ENCODING option encoding.

## Locale Setup Window

In version 8.2, the Locale Setup Window (LSW) has been redesigned to work in conjunction with the new system options described above. Unless you are in Administrator Mode, the LSW will only list the languages that are supported by the current encoding. This allows you to set additional options or safely change to a locale that is supported for that session.

When a new locale is set, the DFLANG system option is set to a value that corresponds to the chosen locale or "English" if no corresponding value is available. The hex value of the Euro character is also set for the locale. The DFLANG and Euro values are stored in the SAS registry.

The TRANTAB option string is stored in the registry with the new setting. As in Version 8.1, the LSW does copy the trantabs from the LOCALE catalog into the SASUSER.PROFILE and SASCFG.HOST catalogs if you have the proper permissions. Note: Please consider SAS Note SN-005169 and SAS Note SN-005171 in this context.

The LSW will also create trantabs using the TRABASE naming convention at your request.

On EBCDIC platforms, the scanner trantab is set in the sixth slot of the TRANTAB option. The TRANTAB option is set when the window is closed.

The LSW does not set the encoding for the SAS session. Please see Locale Setup Window documentation in the SAS System Help for complete information.

The LSW extends the support of the locale in your SAS System. The LSW sets the Euro character that matches the encoding and copies the host-to-host trantabs into place. The host-to-host trantabs have a different purpose than then transport trantabs set in the TRANTAB system option.

The host-to-host trantabs that the LSW sets up are used by PROCs UPLOAD and DOWNLOAD for SAS data set transfer, Cross Environment Data Access (or CEDA), and Remote Library Services (RLS). Please see the *SAS/Connect User's Guide* for information on the procedures. Also, see the *SAS/Connect User's Guide* and *SAS/Share User's Guide* for documentation on RLS.

If users at your site depend on using trantabs that were created by the Version 6 TRABASE sample program, the LSW can be used to copy those trantabs into place as well.

If you are running SAS in NLS Compatibility Mode on an EBCDIC platform, the LSW will select the Traditional EBCDIC mode by default. The LSW also sets the scanner trantab in the sixth slot of the TRANTAB option on EBCDIC platforms. No other trantab slots will be set, so this setting will not override the transport and character classification trantabs set by the ENCODING option.

### **ENCODING= on the FILE, INFILE and FILENAME statements**

The ENCODING= option on the FILENAME, INFILE, or FILE statements allows you to tell SAS what encoding to use to read or write external files. ENCODING= is also available on the INCLUDE and FILE display manager commands.

Use the ENCODING= option when the encoding of the file does not match the encoding SAS uses to process character data. Refer to the section that describes the NLSCOMPATMODE option to determine what that encoding will be. For example, if you are running SAS in NLS compatibility mode, SAS will assume that your external files are in the compiler, or US English, encoding. If the files are in a different encoding, use the ENCODING= option to specify the name of that encoding.

### **How to configure your system**

If you would like to configure your SAS session for a locale other than the default, you have several options. This section describes how you would use the options documented above to get the results you want from the SAS System.

### **Running SAS in NLS Compatibility Mode**

If you are going to run existing SAS applications that were created for a non-English SAS environment, you can set the LOCALE option and run in NLS compatibility mode, or NLSCOMPATMODE. NLS Compatibility Mode sets up your environment so you get results that are compatible with Version 8.1. Most of your Version 6 and Version 8 SAS programs will run without changes. If you are running SAS on an EBCDIC platform, you will also be able to use the characters on your keyboard for SAS syntax instead of making the character substitutions you were required to make in Version 6 and Version 8. For example, if a Danish user running SAS on OS/390 selects a locale of Danish and NLS compatibility mode, that user can use the '\$' instead of substituting the 'Å'.

The LOCALE system option can be set in your configuration file. To run in NLS Compatibility Mode, use the option NLSCOMPATMODE. You can find a list of locale values in the SAS System Help for the host where you run the SAS System.

When LOCALE is set, the ENCODING system option will be set to an encoding that supports the language for the locale. The SAS System expects user data to be in the encoding that matches the ENCODING option. If you prefer an encoding other than the most common encoding for the locale, the ENCODING system option can be set on the SAS command line or in your configuration file. However, if you are running SAS on an EBCDIC platform, you will have to use a SASHELP library that matches the ENCODING option encoding you set. Check with your system administrator before changing the ENCODING option for your EBCDIC platform.

If you are running on an EBCDIC platform, the encoding will be an EBCDIC encoding rather than the corresponding Open Edition encoding. The EBCDIC and Open Edition encodings are based on the same encoding. However, the EBCDIC encodings have a different character used as the NEW LINE.

In compatibility mode, the encoding set by the ENCODING system option will also be used by applications that create output in or establish communications with applications whose syntax or protocols are not determined by SAS. For example, when ODS generates HTML, RTF, and/or JavaScript, by default the output will use the encoding set by the ENCODING system option.

When the ENCODING option is set, the TRANTAB option will always be set to match the ENCODING system option. The transport format trantabs set by the TRANTAB option are used by PROCs CPORT and CIMPORT to transfer SAS data files. They are also used by PROCs UPLOAD and DOWNLOAD for transferring files and catalogs, rsubmitting code to the server, and returning logs and listings to the client. However, the transport format trantabs are not used for SAS data set transfer. Please see the next section for a description of the host-to-host trantabs that are set up by the LSW.

For more information, see the *SAS Procedures Guide* in the Base SAS software for documentation about PROCs CPORT and CIMPORT. Please see the *SAS/Connect User's Guide* for documentation on PROCs UPLOAD and DOWNLOAD.

### **Experimenting with NLS Compatibility Mode Turned Off**

Running SAS with NLS compatibility mode turned off is experimental for version 8.2. If you are interested in creating new SAS applications and you would like to experiment with running SAS outside of NLS Compatibility Mode, the NONLSCOMPATMODE and LOCALE system options can be used to configure your SAS session. You can find a list of locale values in the SAS System Help for the host where you run the SAS System.

When LOCALE is set, the ENCODING system option will be set to the encoding that is most common for the locale on that platform. The SAS System expects SAS syntax, user data and external files to be in the encoding that matches the ENCODING system option. If you prefer an encoding other than the locale encoding, set the ENCODING system option on the SAS command line or in the configuration file.

If you are running SAS on an EBCDIC platform, you will have to use your own characters for SAS syntax instead of making the character substitutions you were required to make in Version 6 and Version 8. For example, if a Danish user running SAS on OS/390 selects a locale of *Danish* and NLS Compatibility Mode turned off, that user should use '\$' instead of substituting the 'Å'.

Also on EBCDIC platforms, the ENCODING will be the Open Edition version of the EBCDIC encoding. Open Edition encodings are identical to the traditional EBCDIC encodings, except that they use the '15'x code position for the NEWLINE character. You will also need to use the SASHELP that has been prepared for the encoding you select.

The Output Delivery System (ODS) will create output using the encoding that matches the ENCODING system option. See the documentation for the Output Delivery System if you would like your output created using a different encoding.

When the ENCODING option is set, the TRANTAB option will be set to match the ENCODING system option. The transport format trantabs that are set by the TRANTAB option are used by PROCs CPORT and CIMPORT to transfer SAS data files. They are also used by PROCs UPLOAD and DOWNLOAD for transferring files and catalogs, rsubmitting code to the server, and returning logs and listings to the client. For more information, see the *SAS Procedures Guide* in the Base SAS software for documentation about PROCs CPORT and CIMPORT. Please see the *SAS/Connect User's Guide* for documentation on PROCs UPLOAD and DOWNLOAD.

Most of the behavior of the LSW does not change if you are running SAS outside of NLS Compatibility Mode. One difference you will see is that the Open Edition Mode is set rather than Traditional EBCDIC.

If you are running SAS on an EBCDIC platform, the LSW will also set the scanner character classification trantab in the eighth slot of the TRANTAB option. No other trantab slots will be set, so this setting will not override the transport and other character classification trantabs set by the ENCODING option.

## Unicode support

Support for Unicode character encoding formats and informats has also been improved in 8.2.

There are four new Unicode formats and informats:

- \$UTF8Xw. reads/writes character data in 8-bit UCS2 (Unicode).
- \$UCS2Bw. reads/writes a SAS value in big endian 16-bit UCS2 Unicode without a byte-order mark.
- \$UCS2Lw. reads/writes a SAS value in little endian 16-bit UCS2 Unicode without a byte-order mark.
- \$UCS2Xw. reads/writes 16-bit UCS2 Unicode without a byte-order mark.

## Examples

**Note:** This example assumes that your session encoding is ISO Latin 1. Your actual session encoding depends on your operating environment and the setting of the LOCALE or ENCODING options.

```
put x $UCS2B.;
```

<i>\$UCS2B. Format</i>		
ISO Latin 1	SAS Internal Representation in ISO	Output characters in UCS-2 Big Endian Unicode (h

(session) value	Latin (hex)	
Särkkinen Björn	53E4726B6B696E656E20406AF6726E	005300E40072006B006B0069006E0065006E0020004
Jérôme Frémot	4AE972F46D6520204672E96D696F74	004A00E9007200F4006D0065002000200046007200E
Åse Østergaard Kjærsgaard	C5736520 D87374657267616172642020 4B6AE672736761617264	00C500730065002000D8007300740065007200670061
Günter Groß	47FC6E746572202047726FDF	004700FC006E0074006500720020002000470072006F

**Note:** The following example also assumes that your session encoding is ISO Latin 1. Your actual session encoding depends on your operating environment and the setting of the LOCALE or ENCODING options.

```
data_null_;
  set temp;
  file datout;
  put x $utf8x30.;
```

<i>\$UTF8X. Format</i>		
ISO Latin 1 (session) value	x characters in ISO Latin (hex)	(x)hex value, \$UTF8X
Särkkinen Björn	53E4726B6B696E656E20426AF6726E2020202020	53C3A4726B6B696E656E20426AF6726E2020202020
Jérôme Frémot	4AE972F46D6520204672E96D696F742020202020	4AC3A972C3B46D6520204672E96D696F742020202020
Åse Østergaard Kjærsgaard	C5736520D873746572676161726420204B6AE67273676161726420202020	C385736520C398D873746572676161726420204B6AE67273676161726420202020
Günter Groß	47FC6E746572202047726FDF20202020202020	47C3BC6E746572202047726FDF20202020202020

## Details

The \$UTF8Xw. format writes a SAS variable in the UTF-8 encoding. The output characters may be one to four bytes, so you must specify a width wide enough to include all of the characters in the variable. The width of the characters will be dependent on the code point value of the individual characters.

You can also specify UTF-8 with the ENCODING= option on the file, infile, and ODS OUT statements. This means you can write UTF-8 output from a SAS session (by converting data on the way out from session encoding to UTF-8), and read UTF-8 encoded external files (by converting data on the way in from UTF-8 to session encoding).

## **Additional Notes**

### **Traditional EBCDIC vs. Open Edition**

On EBCDIC platforms, the newline character (NL) is encoded as '15'x. ASCII platforms use the line feed character (LF) encoded at '0A'x. In SAS, we have created Open Edition encodings, which ensure that the mapping between EBCDIC NL and ASCII LF is correct. The ENCODING option is set to an Open Edition version of the EBCDIC encoding when you are run SAS with NLS Compatibility Mode turned off (NONLSCOMPATMODE).

SAS also supports Traditional EBCDIC encodings that map the NL to ASCII NL, which is position '85'x. The EBCDIC line feed (LF) at '25'x is mapped to ASCII LF at '0A'x. The ENCODING option is set to the traditional EBCDIC encoding for the locale for NLS Compatibility Mode (NLSCOMPATMODE).

### **Locale Setup on the Remote Server**

If you are using SAS/Connect to connect to a remote SAS server, you will need to set up the server session for the locale SAS is using. You must do this after signon to the remote session.

To set up the locale on the remote session, you have two choices:

1. You can run the Locale Setup Window, select your language, and set the Remote Submit option before you close the window. This will upload the proper trantabs to the remote session and, if necessary, will also set the trantab option.
2. You can also achieve the remote locale setup by running the %lswbatch() macro and setting the locale= parameter to match your locale option and the remote= parameter to on. For example, if you are running MVS using LOCALE=Danish, you would signon to your server, then run the macro as follows:

```
%lswbatch( locale=danish, remote=on );
```

If you use an autoexec file to do your signon, you can place the %lswbatch() macro setting in an autocall after your signon statement.