

Translation tables: myth and reality.

Philippe Jalbert, Cogiforce inc, Montreal, QC, Canada.
Normand Létourneau, Cogiforce inc, Montreal, QC, Canada.
Dominic Roy, Cogiforce inc, Quebec, QC, Canada.
Jean Hardy, Cogiforce inc, Quebec, QC, Canada.

Key Words: translation, national language, TRANTAB, codepage, multilingual, remote library, Client/Server, file transfer, data warehouse.

Abstract.

In the always growing world-wide communication, ease of access to numerous databases and environments in various national languages is now fact of life. Correctness of interpretation of language specific characters is becoming a common issue in Client/Server.

This paper is addressing the special characters' translation problem and suggests the **SAS®** Translation Products as a proven effective solution to this problem. The authors are sharing their factual experience with the audience. They are demonstrating with an actual case how complex it may be and how easy it may become. In fact, they will explain why reading a file from outer source may end up at displaying many dark squares instead of readable text.

The authors are not pretending to answer all the questions that may be raised in the translation world, but will guide audience toward the correct path within the SAS solution.

Intended audience: SAS programmer analysts, project leaders, application developers and end users. People dealing with Client/Server, multi-platform and Internet file transfer.

Myth

There are many myths concerning the file transfer and Translation Tables.

Here are some of the mostly heard from people we have talked about this subject.

Myth #1: I will never use Translation Tables because I am using only one language and it is plain English.

Myth #2: I do not use special character between different platforms, so I do not need Translation Tables.

Myth #3: Using Translation Tables need a special training because it is a very cumbersome and complex procedure. We are never sure of what is in there. So, I do not use it.

Myth #4: I avoid using databases, files or variables with special characters because I will not be able to see it correctly.

Myth #5: Using Translation Tables is a breeze because it is fully automated.

Reality

The truth is, total transparency of the platforms and languages is not that evident but, once correctly installed and well tuned between platforms, total translation may become an easy and simple task for the user.

From a well-planned installation and customization of the Translation Tables, anyone may get every translation needed from it.

SAS: the perfect translator.

Why with **SAS**.

Because is a complete integrated portable solution, we believe that SAS is the best Client/Server software. From the experience gained at translating from a platform to another, we found that using TRANSLATE function. TRANTAB procedure and TRANTAB option, was a complete solution. With it, we reached a total control of translation no matter from where to where, and to what type of file, database or variable it was applied to.

Translation may be done on either source or target platforms. Usage of the SAS translation tables is easy and almost maintenance free.

What is what?

TRANTAB procedure.

It is a SAS/BASE procedure used to create or edit **Translation Tables**. There is two types of tables that may be produced by this procedure. The first one is the tables used through the TRANTAB option and activated by the user for a specific translation. The second one is the system modified engine operated tables. These tables are unknown by the user but mandatory to the operating systems to perform a correct translation in a remote library.

Translation Tables built in SAS may translate SAS tables, SAS catalogs, SAS binary transport files and external text files. These files, or catalogs, may contain standard and non-standard characters. The translation tables ensure the integrity of characters' conversion from the originating codepage to the destination one during the file transfer process.

TRANTAB option.

It is a SAS/BASE option used to correctly translate characters during one of the CIMPORT, CPORT, UPLOAD or DOWNLOAD procedure. It is also the option that correctly produce a text-string operation such as lower- and uppercasing, classifying or document scanning.

TRANSLATE function.

The TRANSLATE function is a SAS/BASE and SCL function. It may be used in addition to translation tables, to convert SAS tables or external files between environments. It may also be used to search and replace special characters in a string like a source program, a SAS dataset or an external file. The way to easily use it , is through a macro developed to answer users' specific needs.

Usage of the Translation products.

Translation products may be used to transfer data files between different environments, or between similar environments using different "Codepage". Specific translation tables are mandatory when the transferred files are containing non standard characters.

The non standard characters are known as characters with an ASCII decimal number above 127. A non standard character is a special or accentuated character as ë, î, Û, æ, ç, ñ, ß, £ and ¥. It may also be a border type of character such as ■, ¶, †, ‡ or †.

Who may use Translation products?

Actually, everybody uses translation tables. Almost anybody may use modified translation tables, no matters if your native language is English, French, Spanish, German or else. If you are using files containing any non standard characters, you may benefit from using translation tables.

The typical user is the person who has to transfer data between different environments like the program loading from remote platform as in a Client/Server or datawarehouse environment and that these programs are containing non standard characters.

The occasional user would be someone transferring files between similar platforms loaded with different codepage. This is not considered as a day to day activity.

General comments.

- SAS is the perfect software to link many platforms because translation tables may easily insert where translation need arise. Translation process could be compared to an optical filter. It does not change the essence of the object or character, it only modifies the displaying of it to make it readable to the appropriate platform.
- Translation tables allow usage of many languages in a multi-environment across platforms. Multi-environment applications are commonly using standard characters which avoid most translation errors. Default translations used are then those supplied with protocols that use only standard characters. Yes, **even if you are not aware of it, you are using translation tables in your daily work.**
- A good usage of the translation tables would lead to a better quality of the delivered product by adding national language and borders to it.
- Translation table could be used to eliminate the language barrier to give control to translation between different codepages. This would eliminate the problem of DBMS incompatibility due to presence of non standard characters in these databases.

Technicalities

Multiple languages.

Multiple languages would be made easier by implementing multilingual codepage such as ACSII-850, EBCDIC international CP500 or ISO-8859-1 (ISO-Latin-1) in each of the linked environments. These codepages already contain the commonly used non standard characters for the languages using Latin-based character set. The goal is to simplify the translation process.

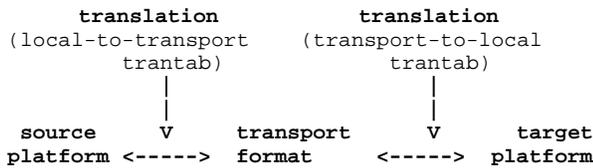
Multiple platforms.

Information in a Client/Server environment is located on different platforms and databases are stored using different DBMS formats (ORACLE, DB2, SYBASE, ...). It also contains special characters associated to any specific language.

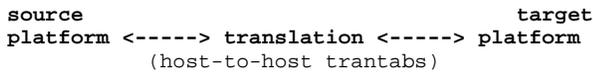
TRANTAB usage.

Kiefer and Kohl (1995) described the technicality of the translation tables' usage in two simple diagrams. We reproduce them below along with our comments.

TRANTAB option is used to correctly execute UPLOAD, DOWNLOAD, CPORT and CIMPORT procedures when a non standard character set is implied in either the source or the target file. This is known as the **across hosts' translation tables**. Refer to the "What is what" section above for more detail.



The **host-to-host translation tables** do not use the TRANTAB option. They are used for any kind of remote library and printing facilities. To correctly interpret non standard characters, SAS supplied translation table names must be used. The modified version of these translation tables must be located within either the SASUSER.PROFILE catalog or the SASHELP.HOST catalog.



In the light of these diagrams, we may say that the translation products allow:

- Translation of the non standard characters associated with foreign languages using Latin type alphabet, however is more complex and usually not automated. Specific translation tables are requested to allow sharing files between environments.
- Translation between similar platforms that are using different codepages, of non standard characters found in foreign language alphabets, like French, Spanish or German.

The content of this paper has been tested and verified in version 6.10 and 6.11 of SAS on PC and UNIX platforms, and version 6.08 of SAS on the mainframe.

Benefits for my organization.

Client/Server application, like datawarehouse, may now use any database containing standard and non standard characters. Translation tables will eliminate any error problem related to the character translation process.

Because every organization is different from one another, we will let you think about any possible implementation of the Translation products that your organization could benefit from.

You may look at the “Experienced installation” section later in this paper, for a good example of TRANTAB and Translation Tables usage. This is a Data Warehousing based installation but this is not limited to the only Datawarehouse environment. Any Client/Server or multi-platform type of environment may use these techniques.

Key points for installation.

Each project will have to evaluate pertinence of using specific translation products or not. We recommend a complete and definitive approach.

The installation of translation products benefits from using a well-prepared checklist to have control over the different codepages, emulation softwares and transfer protocols to obtain the perfect synchronization of the translations between environments and the different codepage combinations.

Installation of translation products and their usage do not follow a unique predetermined procedure. If the codepage combination to be used in the translation process covers more than two environments, or do not appear in the existing SAS Institute Inc. TRANTAB programs included within the 6.10 or 6.11 versions of SAS, the development of specific tables might then be needed.

This paper suggests a list of key points to be verified before installation of translation products. This would probably avoid traps and nightmares.

The main key points are:

- List of the environments implied and the codepage used by each one.
- List of the emulators and their codepages (real and emulated ones if any).
- Uniformity of the codepage combination.
- Physical links for file transfer.
- Transfer protocol list (EHLAPPI, TCP/IP,...) other than SAS.
- List of codepage and special character sets.
- List of existing modified SAS translation tables.
- Perform tests (to be developed in relation with the user community).
- DBMS codepage if any.
- Which platform is client and which is server.

Improper translation results.

With the increasing usage and development of the World Wide Web, we are more frequently seeing users experiencing problem of bad translation in their e-mail texts and attached files transferred.

Perfect synchronization of the translation for special characters between the database and the environment is essential. This table shows the character displayed for a specific hexadecimal value when no translation process occurs. The referenced hexadecimal value is based on the ANSI-Windows codepage.

Hexadecimal Code	c4	e4	cb	eb	d6	f6	c0
ANSI-Windows	Ä	ä	Ë	ë	Ö	ö	À
ASCII-437	-	Σ	Ʀ	ø	F	÷	L
ASCII-850	-	ō	Ʀ	Ù	í	÷	L
ISO-8859-1	Ä	ä	Ë	ë	Ö	ö	À
Roman8	á	ō	ù	Š	ø	-	â
EBCDIC International	D	U	ô	Ô	O	6	{

The next table shows which hexadecimal value has to be used to get to the correct displayed character according to different codepages.

Non Standard Characters	Ä	ä	Ë	ë	Ö	ö	À
ANSI-Windows	c4	e4	cb	eb	d6	f6	c0
ASCII-437	8e	84		89	99	94	
ASCII-850	8e	84	d3	89	99	94	b7
ISO-8859-1	c4	e4	cb	eb	d6	f6	c0
Roman8	d8	cc	a5	cd	da	ce	a1
EBCDIC International	63	43	73	53	ec	cc	64

Experienced installation.

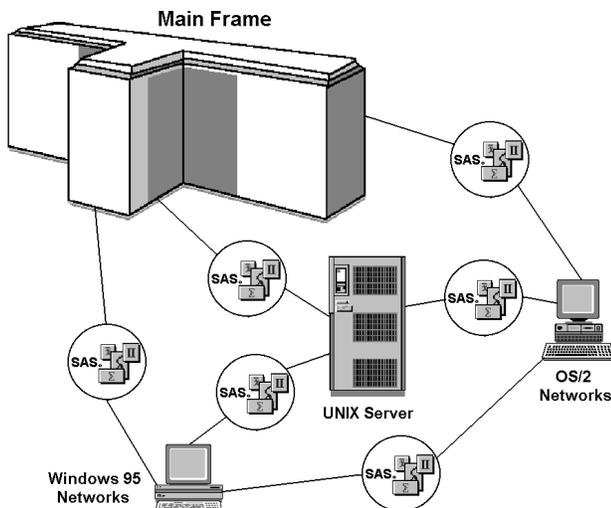
In a major communication company in the province of Quebec, Canada, we have set Translation products to allow them to use any French Canadian character they want to use. This is done in such a way that they may use any terminal, no matters if it is an OS/2 or a Windows 95 based terminal. They may also use it from any station of their network, to any other platform.

The original context where the total translation has been developed at our customer site is this: data is gathered by the mainframe from various sources. According to specific needs, extraction process takes place on the mainframe and is controlled from an OS/2 station. Extracted data is then transferred from the mainframe through the OS/2 network to its final destination; a datawarehouse on a UNIX server.

Their environment is made of a mainframe/MVS/ESA, a HP-UNIX server loaded with SYBASE, some OS/2 servers deserving many OS/2 Warp 3.0 stations, and some Windows-NT servers deserving numerous Windows 95 stations.

All of these platforms are sharing information. Some of these, live, and most of the databases are containing special characters. Translation must then be everywhere and almost transparent to the users' community.

This can be summarized by the figure below, in which we show links between all of the related environments in a single draw.

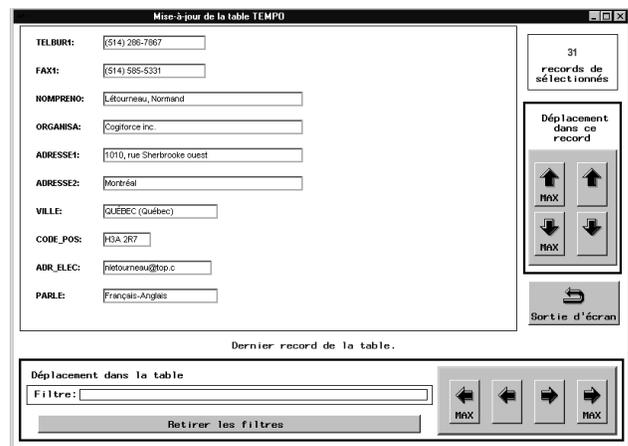


This figure has been simplified to give a good idea of the real connection between the platforms.

Users are querying the Warehouse from either an OS/2 or a Windows 95 station. They may also log to the mainframe to fulfill any specific task. No matter which platform they are looking from, or which platform they are looking at, correctness of displaying is mandatory.

To fulfill this requirement, we have modified some of the SAS supplied Translation Tables to include all of the accentuated characters used by our client. In addition, we have created new TRANTAB tables to accommodate CPORT, CIMPORT and remote library connection.

As a result, users may look at the content of a UNIX database from either an OS/2 or a Windows 95 terminal and see, as shown below, the very same thing.



This figure is a modified content of a SYBASE database. We modified the real content to comply with the client request for confidentiality of the information and structure. The screen itself is the very same as in the real application at our customer location.

On this screen built using Frame Entry of SAS/AF, we may see accentuated characters and they are correctly interpreted no matter what is the used platform for reading. We are also writing to SYBASE using the SYBASE writing standard. Please note that SYBASE may easily be replaced by any commercial DBMS you may prefer to use.

At the origin, this screen and its SCL code was build using SAS on an OS/2 platform. It has been ported to the Windows 95 environment using the above mentioned technique. We also migrated all of the source OS/2 programs using a MACRO written for this environment, namely from OS/2 to Windows 95. A counterpart MACRO is used to port any Windows 95 program to OS/2.

To this, any other platform can be added. We just have to add or modify necessary tables. Once mastered, it becomes a pretty easy task.

Correctness of displayed data is what they get with the SAS products. We do not care at which type of display

device they are displaying their data, even on printers, special characters are perfectly shown and displayed.

Conclusion.

We have to keep in mind that designing and installing specific translation product is not automatic and it may imply complex tasks.

The ultimate goal is to achieve a perfect synchronization of all the SAS translation tables with all the transfer protocols, terminal emulators and codepages that are found in the linked environments.

In other words, make users' life easier by providing the most intuitive environment possible.

Authors' addresses.

Cogiforce inc.

In Montreal
1010, Sherbrooke West, suite 1800
Montreal, Quebec (Canada)
H3A 2R7

Phone: (514) 286-7867

In Quebec City
4715, des Replats, suite 275
Quebec, Quebec (Canada)
G2J 1B8

Internet addresses:

Philippe Jalbert: pjalbert@cogiforce.com
p_jalbert@sympatico.ca

Normand Létourneau: nletourn@cogiforce.com
nletourneau@top.ca

References.

Kiefer, Manfred and John R. Kohl, 1995, "SAS System Support for International Character Sets", *Observations: The Technical Journal for SAS Software Users*, 1995-2. 21-36, Cary. NC: SAS Institute Inc.

Acknowledgments.

The authors would like to thank Mr. Jochen Kristen and Mr. Manfred Kiefer (SAS Institute GmbH) for their expertise and enthusiasm. We used their comments and documentation to build up our own expertise.

SAS, SAS/BASE and SAS/AF are registered trademarks or trademarks of SAS Institute Inc, in the USA and other countries. ® indicates USA registration.