CREATING A SPANISH LANGUAGE WEB PAGE

**Miguel K. Miller - NWSO San Diego, CA
Armando L. Garza - NWSO San Diego, CA**

Introduction

California, the state with the largest number of Hispanics in 1990, also registered the biggest increase of 2.2 million Hispanics between 1990 and 1997. According to latest Census figures released at the end of 1998, Los Angeles County led all counties with an increase of over 649,000 Hispanics over that seven-year period. Other areas in the west that had significant increases of Spanish-speaking residents included Arizona and Nevada. To date, the Hispanic population in California is estimated at 9.9 million, while Arizona has just over 1 million and Nevada is just shy of the 1 million Hispanic population mark.

With these tremendous increases in the Hispanic population, National Weather Service Offices that are located throughout the southern states across the country have been faced with a serious problem of developing outreach programs aimed at this rapidly growing customer base.

With new technological advances and software applications, existing computers are now capable of performing translation functions required to provide Spanish text products to customers over National Weather Service (NWS) web sites, and also provide Spanish language verbal products over the Console Replacement System (CRS). This Technical Attachment will only address translating National Weather Service products from English to Spanish, as is currently in place in San Diego, and describe the process which makes Spanish-language text products available on the Internet.

System Requirements

The task for performing the translation from English to Spanish can be accomplished in a variety of ways. While every office interested in performing this task may take its own unique approach or use a different method to accomplish the translation, the following method is used by the San Diego Weather Forecast Office (WFO) to automatically translate several routine products from English to Spanish.

Offices desiring to run the translation software should use a dedicated PC. Experience has shown that it is not advisable to run other programs on the same PC since this can result in the disruption of the translation programs.

In order to run the translation software, the dedicated computer should meet the following minimum specifications: Pentium 133 PC;   .64MB RAM;   .Windows 98/95;   28 MB hard disk drive space; CD-ROM drive.

Presently, a Procomm Plus scheduler is being used to initiate the translation process. Because the scheduler method is the only way to initiate the translation at this point in time, only routine products are being translated on a time schedule. This also requires that all of the scheduled products be issued in a timely manner. In the future, another program could be set up to recognize when a product is issued, which would in turn initiate the translation. This would be necessary for all non-routine products such as a warning, or non-routine issuances of routine products such as an update to the zone forecast product.

Translation Software

For many years, offices throughout the country have been experimenting with different types of translation software programs. Offices in Puerto Rico, Florida, and Texas, as well as those in Arizona and California which have high Hispanic populations, have worked hard to identify software that would help the National Weather Service disseminate products to these large communities.

During the past two years, a combined effort by the Tucson and San Diego offices has resulted in some success in working with a software translation program which is being used to make products available on the office web sites. The setup requires the expertise of someone with a good working knowledge of the Spanish language who can be assigned to work with several software packages (such as WinBatch). If the individual works on this project during slow periods of his/her operational shift and is given a few administrative type shifts, the process to set this up will take 3-4 months. Several more months experiencing different seasonal phenomena will be spent in working with the dictionaries. Any office should then be able to run the software to produce the web site products within 6-8 months.

WinBatch is a Windows-driven software package available at a very reasonable cost and provides the ability to manipulate applications. In San Diego, the WinBatch system, which allows mouse-driven software to be used, is tasked with executing the translation portion of the program using the WBT language.

Two separate software programs are used to accomplish the translations:

a.   Globalink Power Translator version 6.0. This is used for text-intensive products such as the Zone Forecast Product (ZFP) and the State Forecast Product (SFP).

b.   Tolken97 Translator version 3.2. This is used for numeric-intensive products such as Hourly roundups (SWR) and Daily Temperature and Precipitation Summaries (RTP).

Although the processes are similar for both translators which are used for development of the Spanish web page, they will be discussed separately.

Procedures Using Globalink

The scheduler starts the batch file assigned for each product at a designated time. The batch file (filename.bat) runs several other embedded programs which are described below:

First, using FTP, the batch file searches any computer which stores AFOS or AWIPS products for the desired product and acquires/grabs it.

It then starts another short program called a script program (filename.scr) which deletes the old product and places the new acquired product into the directory containing the batch file.

A third program is run to save the new product in a format which the translator needs for input: the Rich Text Format (RTF). This program (filename.wbt) places the product in a word processor, such as Microsoft Word, and simply saves it as a Rich Text Format file (productname.RTF).

The batch file then starts the fourth program (convfilename.wbt). The WBT language used by Winbatch is designed to execute the following tasks in order: call up Globalink, prepare or "clean up" untranslatable portions of the text, translate the product, save it as a text file (productnamesp.txt), and finally exit Globalink. (See Appendix B)

Dots and other symbols found in NWS products inhibit the translator from making a complete translation, so several "find and replace" commands are used to "clean up" the text.

Once the translation is complete, the batch file starts the fifth program (upfilename.scr) which uploads the file to the web page automatically. Finally, the batch file terminates itself.

This process takes several minutes, depending on the length of the product. An average Zone Forecast Product, for example, will take five to seven minutes from start to finish.

Procedures using Tolken97

As with Globalink, the scheduler starts the batch file associated with the product. Then, as with the previous program, the batch file will run several other commands.

First, it captures the latest recently issued product.

Next, the script file (filename.scr) deletes the old product and inserts the new one. Tolken97 has no trouble translating the product in its current format, so it is not necessary to convert to Rich Text Format (RTF) by using the "productname.wbt" program as with Globalink.

The translation program is then initiated using the "convfilename.wbt" file. There is also no need to "clean up" or alter the text in any way. After a few seconds, the translation is complete and the file is saved as "productname.trs", which is then copied to "productnamesp.txt".

The upload program sends the file to the web page using FTP. The batch file then terminates itself. (See Appendix C)

This process takes less than one minute. The question that might arise is: "If Tolken97 is so fast, and has no selective formatting preferences, why is it not used to translate all the products?" The answer is simply that the Tolken97 dictionary is limited and not as robust as the dictionary in Globalink. Perhaps with more work, the dictionary in Tolken97 can be augmented enough to equal the capability of the dictionary in Globalink.

Editing Dictionaries

After initial installation, the dictionaries for both Globalink and Tolken97 require much work before they can be used for translating weather terms. They do not include even the most simple meteorological terms. Since the National Weather Service uses a very unique and selective language, most software provided dictionaries can be described as being "meteorologically challenged".

In order to correct this deficiency, San Diego began with a small glossary of terms provided by the Weather Service Office in San Juan, Puerto Rico. This was very helpful for many NWS terms, especially marine weather terms such as "Small Craft Advisory", etc.

Since Puerto Rico does not experience the same types of weather phenomena that California does, it was necessary to make manual translations for such phrases like "Winter Storm Warning." These additional translations do not include every possible weather scenario for every office, but they encompass everything that was listed in the San Diego Station Duty Manual.

It should be noted that both translator programs have the ability to translate phrases, or even full sentences, such as a title of a product, e.g., "Extreme Southwestern California Zone Forecast."

Tying into the World Wide Web (WWW)

Several modifications were required in order to prepare translated products for dissemination over the web.

> First, it was necessary to execute the translation software to a particular product. After receiving a completed translation, identification and correction of meteorologically incorrect translations were required. When new weather scenarios occurred, new words and phrases appeared in the products. Over a period of many months, nearly every weather phenomenon was experienced, and the new words and phrases were added for completion of a reasonably complete dictionary. For example, in southern California, the main weather scenarios are winter storms, low clouds and fog, Santa Ana winds, and monsoon thunderstorms.

> Second, the products were placed on the scheduler, translated automatically and manually, checked for problems (but not uploaded to the web page). This also provided an opportunity for troubleshooting bugs or glitches which might exist in the Winbatch programs.

> After a trial period of constant vigilance and improvements, the translations became reasonably acceptable and the programs ran smoothly. At this point, these translated products were uploaded and added to the web page. It is important to not add the translated products to the Internet before such a trial period is completed,

As with any computer files, it is recommended that backup files be created for the dictionaries. A failure by the PC will erase all edits. Edits were saved which are necessary for most National Weather Service translations. Each office, however, will need

to add their own edits, mainly for local geography nomenclature, localized events, or even specific forecaster writing styles.

Life would be much easier for the translator if all forecasters had good grammar skills and similar styles. Very choppy, incomplete or poor English is also used for the sake of brevity. Since these problems exist, creative solutions need to be made to the dictionaries to accommodate these local translation problems. For example, "Partly cloudy in the afternoon" vs. "Partly cloudy afternoon." Both phrases must be accounted for in the dictionary, using the Spanish equivalent for "in the afternoon."

Cautionary Tips and Identified Deficiencies

The following are some deficiencies by both translators which have been identified and have not yet been satisfactorily resolved. Two translations cannot run at the same time as competing Winbatch commands will interfere with each other.

Globalink
- It reads each new line of text as a new sentence. Disjointed translations from one line to the next are often the result, and some are egregious.

- New lines are often begun before a previous line of text is complete, creating a choppy look to the finished product. This effect has some correlation to the problem with interpreting new sentences.

- Translated phrases, a group of words translated as a complete phrase such as "low clouds and fog," will be translated into lowercase. Normally the translated text remains in its original capitalized form. The final translation is a mix of upper and lowercase letters.

Tolken97
- It has difficulty maintaining the alignment of numeric columns. This is apparently caused by the differing lengths of weather phenomena translations.

- Several problem words or abbreviations cannot all be accounted for and translated correctly. In one case, the abbreviation "IN" is used to indicate inches of precipitation. This is translated into the Spanish equivalent of the English word "in."

Concluding Remarks

Most National Weather Service offices outside of San Juan, Puerto Rico have no Spanish-speaking experts who can manually translate products for Spanish-speaking customers.

Even for those offices who have staff members capable of translating, it is not possible for them to translate all products every day. In an effort to make National Weather Service products available to a greater number of customers, an automated system for translating products into Spanish was developed. These translated products are now available through office web pages and NOAA Weather Radio. This new translation automation system is an evolving process, but the basic functions are in place and executing, and have shown much promise.

In San Diego, the web site is now being used by the print media, television, and the general public interested in planning outdoor activities. Using the Internet to access Spanish weather products has been favorably accepted by the customers. It is also a valuable outreach in an area where the Hispanic community is large. The San Diego WFO serves an extremely large Hispanic customer base in California with the highest Hispanic populations residing in Los Angeles County (4.0 million), Orange County (750,000), San Diego County (700,000) and San Bernardino County (530,000). Spanish media and emergency managers in Tijuana, Baja California Mexico also access the web site in their planning and decision-making efforts.

Appendix A provides a step-by-step procedure on how to get started using the translation software programs. Appendix B is an example of how San Diego is using Globalink to translate the zone forecasts. Appendix C is an example of how Tolken97 translates the more numeric intensive Temperature and Precipitation Summaries.

Finally, it should also be noted that the translation software programs are capable of working with other languages, such as French, German, Italian, and Portuguese. Therefore, depending on the locality of the weather offices and the requirements for serving large, non-English speaking customers, this process can be used in a similar fashion to reach those residents.

APPENDIX A

Getting started

1. Dedicate a PC for use in translating of the products. (San Diego is currently using a Pentium 133 with 64MB RAM and 1GB Hard Drive running Windows'95).

2. Install a Procomm Plus scheduler, Globalink Power Translator Version 6.0, Tolken97 Version 3.2, and Winbatch. Become familiar with their use.

    Globalink Incorporated
    1-800-255-5660
    email: info@globalink.com
    Internet: http://www.globalink.com

    Tolken97
    email: hagsten@algonet.se
    Internet: http://www.algonet.se/~hagsten

    Tolken97 is shareware which you can temporarily download and test drive. It is also very affordable for purchase. In March 1999, the cost was $15. This provides a license to the user.

    Winbatch
    1-800-762-8383
    email: wwwtech@halcyon.com
    Internet: http://www.windowware.com

3. Download backup dictionaries into the translator dictionaries and create your own backup.

4. Write programs to download, prepare, translate, and upload the products (see examples in Appendices B and C).

5. Manually initiate numerous translations in a variety of weather scenarios.

6. Edit dictionaries to include local translation needs (a Spanish speaker and English-Spanish dictionary would be useful).

7. Set up Procomm Plus scheduler to automatically run programs.

8. Have a highly motivated individual take responsibility to continually edit dictionaries ensure the smooth running of the translating programs.

APPENDIX B

A Globalink Example Translating the ZFPSGX

First, the zfp.bat file is run. This file begins and ends the entire process. It initiates all other programs to run as soon as the previous task is completed.

```
@echo off
cd c:\download
del c:\download\laxzfpsgx
ftp -n -s:c:\download\zfp.scr
c:\progra~1\winbatch\system\winbatch.exe zfp.wbt
c:\progra~1\winbatch\system\winbatch.exe convzfp.wbt
pause
cd c:\download
ftp -n -s:c:\download\upzfp.scr
```

Second, the zfp.scr file is run. This file retrieves the desired product from a system containing AWIPS or AFOS products. (The ftp address, username and password are uniquely defined by each office).

```
open (ftp address)
user (username password)
ascii
get LAXZFPSGX
disconnect
bye
```

Third, the zfp.wbt file is run. This file puts the product into a wordprocessor, then saves it as a Rich Text Format (RTF) file so Globalink will translate properly. The language is the WBT language which executes commands by using the keyboard instead of the mouse.

```
Run("c:\progra~1\accessories\wordpad.exe","c:\download\LAXZFPSGX")
TimeDelay(5)
SendKey("!f")
TimeDelay(2)
SendKey("a")
TimeDelay(2)
SendKey("zfpsgx.rtf")
SendKey("{TAB}")
SendKey("r")
SendKey("!s")
SendKey("y")
SendKey("!f")
SendKey("x")
```

Fourth, the convzfp.wbt file is run. This file (in WBT language) calls up Globalink, opens up the product to be translated, prepares it to be translated by "cleaning up" the formatting glitches, translates it, then saves it as zfpsgxsp.txt.

```
Run("C:\ptwin60\ptwin60.exe","C:\download\zfpsgx.rtf")
TimeDelay(14)
SendKey("~")
TimeDelay(7)
SendKey("!e")
SendKey("r")
SendKey("...")
SendKey("{TAB}")
SendKey(", ")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("~")
TimeDelay(14)
SendKey("y")
TimeDelay(8)
SendKey("~")
SendKey(".<")
SendKey("{TAB}")
SendKey(" ")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("~")
TimeDelay(5)
SendKey("y")
TimeDelay(10)
SendKey("~")
SendKey(".T")
SendKey("{TAB}")
SendKey("T")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("~")
```

```
TimeDelay(4)
SendKey("y")
TimeDelay(10)
SendKey("~")          ...(There are many other similar "clean up" commands, quite long)...
```

And now, the actual translation:

```
SendKey("^d")
TimeDelay(200)
SendKey("!f")
SendKey("{DOWN}")
SendKey("{DOWN}")
SendKey("{DOWN}")
SendKey("{DOWN}")
SendKey("{DOWN}")
SendKey("{DOWN}")
SendKey("{DOWN}")
SendKey("{DOWN}")
SendKey("~")
TimeDelay(1)
SendKey("zfpsgxsp.txt")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("{DOWN}")
SendKey("{DOWN}")
SendKey("{TAB}")
SendKey("{TAB}")
SendKey("~")
SendKey("y")
TimeDelay(3)
SendKey("!f")
SendKey("x")
SendKey("n")
TimeDelay(5)
SendKey(" ")
TimeDelay(15)
```

Finally, the upzfp.scr file is run. This file uploads the translated product and inserts it into the web page. The ftp address, the username, and the password are uniquely defined by each office.

```
open (ftp address)
user (username password)
ascii
lcd c:\download
cd htdocs/sandiego
put ZFPSGXSP.TXT
disconnect
bye
```

APPENDIX C

A Tolken97 Example Translating the RTPSGX

First, the rtp.bat file is run. This file begins and ends the entire process. It initiates all other programs to run as soon as the previous task is completed.

```
@echo off
cd c:\download
del c:\download\laxrtpsgx
ftp -n -s:c:\download\rtp.scr
c:\progra~1\Winbatch\system\winbatch.exe convrtp.wbt
pause
copy laxrtp~2.trs rtpsgxsp.txt
cd c:\download
ftp -n -s:c:\download\uprtp.scr
```

Second, the rtp.scr file is run. This file retrieves the desired product from a system containing AWIPS or AFOS products. (The ftp address, username, and password are uniquely defined by each office).

```
open (ftp address)
user (username password)
binary
get LAXRTPSGX
disconnect
bye
```

Third, the convrtp.wbt file is run. This file (in WBT language) calls up Tolken97, opens up the product to be translated, translates it, saves it as rtpsgx.trs which is then copied to rtpsgxsp.txt.

```
Run("c:\progra~1\tolken97\tolken97.exe","c:\download\laxrtp~2")
WinActivate("Tolken97")
TimeDelay(7)
SendKey("^{F7}")
SendKey("~")
TimeDelay(12)
SendKey("^a")
TimeDelay(1)
SendKey("!s")
TimeDelay(2)
SendKey("!f")
SendKey("x")
TimeDelay(3)
SendKey(" ")
TimeDelay(8)
```

Finally, the uprtp.scr file is run. This file uploads the translated product and inserts it into the web page. (This is how it is done in San Diego, other offices may vary). The ftp address, username and password are uniquely defined by each office.

```
open (ftp address)
user (username password)
ascii
lcd c:\download
cd htdocs/sandiego
put RTPSGXSP.TXT
disconnect
bye
```