3-18-1988

MicroMATER: a Proposed Standard Format for Microcomputer Terminology Files

Don Chapman

Alan K. Melby

Follow this and additional works at: https://scholarsarchive.byu.edu/dlls

BYU ScholarsArchive Citation

This Article is brought to you for free and open access by the All Journals at BYU ScholarsArchive. It has been accepted for inclusion in Deseret Language and Linguistic Society Symposium by an authorized editor of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
MicroMATER: a Proposed Standard Format
for Microcomputer Terminology Files

Don Chapman and
Alan K. McQuay
Brigham Young University

This paper will first discuss a few reasons why a standard for microcomputer terminology files is needed and then explain the proposed MicroMATER standard. Finally, GLOSnost, a software program developed by the Translation Research Group (TRG) at BYU, will be described.

I. Why a standard is needed

To begin with, three observations should be noted:

Observation 1:

Language is dynamic, and one of the most characteristic features of language change today is the constant addition of new terms to our vocabulary. Because of the rapid technical growth in today's society, new terms are continually coined to describe new inventions, processes, objects, etc. In the last 50 years, for example, terms have been invented to describe everything from space exploration to genetic engineering. Of course most of these terms remain specialized within the field that they grew out of, such as the names of the parts of a clutch for a BMW. Many terms, however, have probably reached general status. Since microcomputers have become a consumer item, for example, many adults have probably acquired such terms as word processor, floppy disk, and database.

Observations 2:

English is not the universal language of the world. While English may be more of a global lingua franca than any other language, the other languages of the world aren't going away. If anything, the idea of national languages is becoming stronger. Just a few years ago, few companies bothered to translate the manuals that accompanied their products. The rationale, apparently, was that if a company in Denmark wanted to use a mainframe computer built in the U.S., it would have to find someone who knew English to run it. Now, most consumers won't be as willing to find someone who understands English to help them with their new computer; they'll just buy a computer from a company that has already translated the instructions. In other words, as technical products become more consumer-oriented, the need to translate the documentation accompanying them becomes greater.

Observation 3:

Companies in the United States must not ignore the consumers of other
countries. The U.S. trade-balance is bad enough, and most would agree that the U.S. should try to even the balance in any way possible. Naturally, American companies stand a better chance of selling products if they make them accessible to the consumer; if that means companies should translate more documents, then let them get busy translating. Imagine how well Toyotas would sell if their dash boards came with only Japanese characters and owner's manuals written only in Japanese.

From these observations, two points about translation should emerge:

1. The need for translation increases as international trade increases.
2. Many of the documents being translated (owner's manuals, technical documentation, etc.) are filled with new and specialized terms.

These characteristics of translation present a couple of problems for the modern translator. First, although a person may be a qualified translator, that person probably doesn't know all the terms used in a technical document. Say, for example, that a competent German-English translator has been hired to translate the parts list of a packaging machine made in Austria. Would she be able to translate that list without help? Probably not. She would probably have to consult a specialist to find out the English equivalents for the part names.

Second, the translator ought to be uniform in the use of a term throughout the document, even though several equivalent terms in the target language may often be available to the translator. If a person is translating a car repair book and calls a particular part a headlight dimmer in chapter 1, then that person should not turn around and call that same part a lantern intensity reducer in chapter 5. It would be confusing. And to add to the problem of keeping terms uniform, a document is often translated by more than one person, each working on a separate part. In this case it becomes crucial that all translators have access to the translation equivalents which have been agreed upon for key words.

These problems underscore the importance of managing terminology. Previously, terminology management has been largely a tedious chore for translators, which has involved consulting with a specialist to come up with equivalents for key technical terms, maintaining a file of terms, and perhaps even creating a glossary. Including several translators on a project has compounded the problems of terminology management since any changes to one glossary must be made accessible to all translators on that project. Understandably, managing terminology with manual methods, such as using 3 x 5 cards and creating paper glossaries, proves bothersome to translators. Most translators don't like to interrupt their work to shuffle papers or look up a term in a glossary. Electronic terminology managers, on the other hand, make these tasks much simpler. A translator merely needs to gather his terms and store them in a computer file. The terminology manager can then keep track of the terms, and when it comes time to look up a word, the translator simply types a few keystrokes, without even leaving the word processor.

Currently three such terminology managers exist: ABCWord (tm) from ALP Systems, Term Tracer (tm) from INK International, and Mercury/Termex (tm) from LinguaTech International. These terminology managers certainly
make it easier to maintain uniform translations of new and specialized terms. However, the presence of three different terminology managers brings up the problem of compatibility among the proprietary formats of the terminology managers. A glossary created on Termex (tm), for example, cannot be accessed on ABCWord (tm), nor can a glossary created on ABCWord (tm) be accessed on Term Tracer (tm). So what is to be done? Should all translators agree to use the same terminology manager so they can easily exchange files? Perhaps, but a more reasonable solution is to adopt standard conventions for terminology files so that those files can be exchanged among any terminology managers. With the existence of a standard, the companies that produce these terminology managers can readily accommodate their clients by writing one set of conversion routines that changes their proprietary format to the standard format and vice versa. Thus a translator could create a glossary using Termex (tm), convert it to a standard exchange format, and give it to a friend who uses Term Tracer (tm). That friend could then convert the glossary from the standard exchange format to Term Tracer (tm) format. In this paper, MicroMATER is proposed as that standard exchange format.

II. What is MicroMATER

The International Standards Organization (ISO) has already defined a standard called MATER for exchanging information between mainframe term banks via industry-standard 9-track 1/2 inch tapes. MATER is an important step and means that each term bank need only write one import/export utility that converts between its proprietary format and the MATER format. Given the explosive growth of the use and power of microcomputers over the past few years, the next logical step is a companion standard to MATER suitable for microcomputers as well as mainframes. This is the MicroMATER standard the BYU-TRG is proposing. (The MATER standard can be obtained from ISO by specifying that it came from ISO/TC 37 and is standard number 6156.)

The MicroMATER standard consists of five conventions. A file conforms to the MicroMATER standard if it conforms to these conventions which are grouped together as follows:

1. General conventions.
   a. Each MicroMATER file must consist of any number of lines, each ending with an EOL character (CR/LF or LF, depending on the operating system).
   b. A soft return in MicroMATER is represented by the sequence "@;EOL" (at-sign followed by a semi-colon followed by an End of line character). The soft return allows lines to be shortened without losing information. One reason for using soft-return is that many text editors and telecommunications programs allow only 70 characters per line. When using these programs, a person can insert a soft-return after 65 characters, and then reconstitute the lines to their full length for programs that don't specify limits for line lengths. (It is suggested that lines be no more than 65 characters long for maximum portability).
   c. The '@' character (at-sign) is a literal signal, and a character following it is taken as a literal character (except for the ';', which is taken as a soft-return). A literal at-sign, then, is
2. Format conventions.

The MicroMATER format conventions essentially provide the structure of a glossary by defining beginnings of entries (which are varying length records), ends of entries, beginnings of terms, ends of terms, and other aspects of the structure. Briefly, the Format Conventions consist of a header at the beginning of a glossary, any number of records delimited by the Format Marks (which aren't given here, but can be found in the MicroMATER definition accompanying GLOSnost), and the MicroMATER end of file mark.

3. Term Conventions.

a. Each term must be unique. When two terms are identical, it is suggested that uniqueness be enforced by modifying one of the terms in some way, such as appending two spaces and some distinguishing label. For example, "term [part 2]" is made distinct from "term" by adding two spaces and the label "[part 2]."

b. A term may be any length, but it is suggested that a term be no more than 60 characters long. If a term exceeds 60 characters, it is suggested that the term be truncated to 60 characters and the remaining characters be placed as the first field of that entry.

4. Field Names.

Glossary entries often give several pieces of information about a term, such as usage notes, grammar notes, and contextual examples. In MicroMATER, each separate piece of information is represented as a field. The names of these fields are given between curly braces such as "{2}". The standard field names haven't been established firmly, and are open for debate (Infoterm and others have proposed various naming systems), but provisionally, these suggestions apply:

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Field Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>abbreviation or acronym</td>
</tr>
<tr>
<td>c</td>
<td>contextual example</td>
</tr>
<tr>
<td>t</td>
<td>translation of contextual example</td>
</tr>
<tr>
<td>d</td>
<td>definition</td>
</tr>
<tr>
<td>f</td>
<td>field (ie subject matter, such as chemical engineering, biology etc.)</td>
</tr>
<tr>
<td>g</td>
<td>grammar note (e.g. gender, number, case, tense, etc.)</td>
</tr>
<tr>
<td>n</td>
<td>note (some piece of information that does not seem to fit in any of the other categories.)</td>
</tr>
<tr>
<td>o</td>
<td>opposite or antonym</td>
</tr>
<tr>
<td>p</td>
<td>preferred (i.e., a given translation is preferred by a specific client, etc.; it would be followed by the client's name).</td>
</tr>
<tr>
<td>s</td>
<td>source of the term (i.e., a particular reference, document, person, etc).</td>
</tr>
<tr>
<td>u</td>
<td>usage note (e.g. Country X only)</td>
</tr>
<tr>
<td></td>
<td>cross reference</td>
</tr>
</tbody>
</table>
Any number not accompanied by one of these abbreviations is a field name for a translation equivalent (e.g. "{1}"). If a number precedes one of these abbreviations, then the abbreviation refers to the translation equivalent that shares the number. For example, ",{1g}" is a grammar field that gives information about translation equivalent "{1}". If the abbreviation is by itself, for example "{g}", then that field refers to the entire term.

5. Code Conventions.
The European languages use several characters that aren’t used in English, such as the u-umlaut or the c-cedilla. The representations of these characters among micro-computers and various software programs such as word processors is hardly uniform. For that reason, MicroMATER defines code conventions for representing the non-English characters. These codes consist of accent letter codes for representing accented letters and punctuation codes for representing punctuation marks, such as the upside-down question mark in Spanish.

These accent letter codes and punctuation codes are represented by two characters. The first character (called a signal) more or less modifies the second character. For example, in represents the Spanish enyay. The character codes are as mnemonic as possible—the a suggests an a-grave and the two circles in the percent sign suggest an umlaut in %u.

In addition, any character signal, (%,/\,etc) can be given a literal representation by preceding it with an at-sign. Thus, there is no ambiguity in representing accented and punctuation characters.

There has been some work towards creating a standard for representing non-Roman character sets, such as Arabic and Cyrillic. In the future, MicroMATER will incorporate this work and provide a standard definition for several character sets.

To summarize, the five MicroMATER conventions consist of:
1. General Conventions.
2. Format Conventions.
3. Term Conventions.
4. Field Name Conventions.
5. Character Conventions.

The benefit of dividing the MicroMATER standard into these conventions is that a document may conform to some of the conventions, but not others, and still be manipulated by software tools such as GLOSnost. For example, a person may want to send a letter to a friend via a network. Of course that person won’t want to make the letter a glossary (i.e. include a header, records and fields), but he or she may still need to use accented characters. To do so, the person can make the letter conform to the Character Conventions, and then use GLOSnost to adjust the line sizes of the letter and prepare it for sending across telecommunications channels.

III. What is GLOSnost

GLOSnost is a play on the term glasnost made popular by Gorbachev and is intended to suggest openness in the exchange and manipulation of glossary
information. The current version of GLOSnost was written in the C programming language by the TRG and runs under the PC-DOS operating system. It is the intent of the TRG to enhance GLOSnost considerably over the next few years and to make it available under several popular operating systems. In keeping with its name, GLOSnost is a software program which, although copyrighted by Brigham Young University, can be freely copied and distributed so long as the University is not held responsible for results of any problems in the software. Please send any enquiries concerning GLOSnost to

Alan K. Melby
Linguistics Department
2429 JKLB Brigham Young University
Provo, UT 84602.

GLOSnost is used to manipulate files that conform to the MicroMATER standard. The TRG will make no outlandish claims about this program--GLOSnost doesn't clean rooms, make beds, or regulate the oxygen level in houses. Instead it is a collection of utilities. Yet, even if some of these utilities are pedestrian and devoid of glamour, they can come in surprisingly handy for those who work with MicroMATER files or who share files with other systems. For example, the first function, SANITIZE, is useful when a person receives files from another person, via telecommunications networks or otherwise. Often such files contain control characters--perhaps from garble of the network transmission, or perhaps from the word processor that the sender used, or perhaps from something else. These control characters often foul up word processors, so until the control characters are removed, the file is inaccessible. SANITIZE provides a tool to remove those irksome control characters.

Probably the most exciting function in GLOSnost currently is the REVERSER which reverses the language directions of a file that conforms to the MicroMATER standard. For example, a German to English glossary can be changed into an English to German glossary by merely running it through the REVERSER.

The other functions in GLOSnost convert MicroMATER character codes to IBM PC characters and vice-versa, and verify whether a file is in MicroMATER format.

In the future, several more features will be added to GLOSnost, such as
1. Conversion to and from Dbase III (tm) format;
2. Specialized sorter;
3. Secondary key generator;
4. Duplicate term processing;
5. Glossary comparison;
6. Term extraction

IV. Conclusion

It certainly is time to adopt a standard in microcomputer terminology files. Nearly everyone will benefit from having a standard. Translators will benefit since it will be easier to exchange terminology files. Companies
that produce terminology managers will benefit, since they will only have to write one set of conversion routines to and from the MicroMATER standard. The BYU TRG hopes that MicroMATER catches on and that GLOSnost can help persuade people to use MicroMATER conventions.
Appendix

SAMPLE GLOSSARY

Glossary: MANENG2

Language(s): German-English

-------------------

4 Adapter
{1} adapter
{1} p.s.
{d} electronic adapter used to test circuits

4 %ndern
{1} change
{n} make a change in something, such as changing a dimension; see also wechseln
{1} p

4 Andruckrolle
{(a, b)} Haspelandruckrolle
{1} p

anfahren
{1} drive
{1} press control system
{1} press safety materials
{c} Die beiden Endschalter sind so zu montieren, daß sie von
getrennten Kurvenscheiben angefahren werden.
The two limit switches should be assembled in such a way that
they are driven by separate cam plates.{1} p

4 angelassen
{1} hardened and tempered
{a} as {1} p

4 anhebbare Rollen
{1} lifter rollers
{c} Zur leichten Verfahrenheit sind anhebbare Rollen
vorgesehen auf dem Wagentisch. {s} Weingarten materials
{1} p

4 Anheberollen
{1} lifter rolls
{c} Anheberollen im Tisch a.d. Schnittstufe und Umformstufen {s}
Weingarten invoice {1} p

-------------------
adapter

{1} Adapter
{1f} p.s.
{1d} electronic adapter used to test circuits

change

{1} %ndern
{1n} make a change in something, such as changing a dimension; @; see also wechseln
{1f} p

drive

{1} anfahren
{1f} press control system
{1s} press safety materials
{1c} Die beiden Endschalter sind so zu montieren, daß sie von @; getrennten Kurvenscheiben angefahrene werden. The two limit @; switches should be assembled in such a way that they are @; driven by separate cam plates.
{1f} p

hardened and tempered

{1} angelassen
{1a} as
{1f} p

lifter rollers

{1} anhebbare Rollen
{1cG} Zur leichten Verfahrbarkeit sind anhebbare Rollen @; vorgesehen auf dem Wagentisch.
{1s} Weingarten materials
{1f} p

lifter rolls

{1} Anheberollen
{1c} Anheberollen im Tisch a.d. Schnittstufe und Umformstufen
{1s} Weingarten invoice
{1f} p
GLOSnost Menu

GLOSnost MicroMATER Utilities
Version 1.0
(C) 1988 BYU Translation Research Group

S) Sanitize a file (convert control characters to question marks
A) Convert from IBM (8-bit) characters to microMATER (7-bit) ASCII codes
I) Convert from microMATER (7-bit) ASCII codes to IBM (8-bit) characters
R) Reverse language direction of a glossary in MicroMATER format
V) Verify file for correct MicroMATER conventions
X) Exit