AN EXPERIMENT IN BLIND TERMINOLOGY INTERCHANGE: DEVELOPING AND

TESTING CONVERSION ALGORITHMS FOR EXTERNALLY SUPPLIED DATA

by

Aaron D. Alder

A thesis submitted to the faculty of

Brigham Young University

In partial fulfillment of the requirements for the degree of

Master of Arts

Department of Linguistics

Brigham Young University

December 1998

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Aaron D. Alder

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

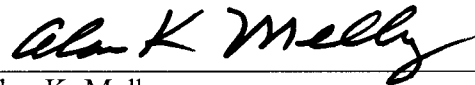|	
Dec. 16, 1998	*Alan K. Melby*
Date	Alan K. Melby, Chair
Dec 16, 1998	*John S. Robertson*
Date	John S. Robertson
Dec 16, 1998	*Deryle Lonsdale*
Date	Deryle Lonsdale

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Aaron D. Alder in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Dec. 16, 1998
_____
Date

Alan K. Melby
_____
Alan K. Melby
Chair, Graduate Committee

Accepted for the Department

John S. Robertson
Department Chair

Accepted for the College

Van C. Gessell
Dean, College of Humanities

ABSTRACT

AN EXPERIMENT IN BLIND TERMINOLOGY INTERCHANGE: DEVELOPING AND
TESTING CONVERSION ALGORITHMS FOR EXTERNALLY SUPPLIED DATA

Aaron D. Alder

Department of Linguistics

Master of Arts

MARTIF (MAchine-Readable Terminology Interchange Format) is an SGML-based format for data interchange among concept-oriented terminological databases. MARTIF is intended for interchange between partners (e.g., two translation companies) who know about each other and are able to "negotiate" details of the format to minimize information loss.

A contrasting approach to terminology interchange is called "blind" interchange. With this approach, details are pre-defined so that export and import routines can be written without "seeing" the interchange partner. However, blind interchange may hinder preservation of information. Complete preservation and total blindness appear to be incompatible. An acceptable blind interchange format must necessarily provide adequate preservation of information.

Hardman (1996) proposed a format for blind interchange based on a 1995 draft version of MARTIF. The present experiment tests a revised proposal for blind interchange

that is based on the 1997 draft version of MARTIF. This revised Blind MARTIF proposal, which is a restricted subset of MARTIF, benefited from comments on the content of the Hardman thesis.

The thesis of this work is that the current version of Blind MARTIF is capable of sufficient preservation of information when blindness is imposed. This work begins with a brief motivation for interchange and a discussion of the tension between negotiated and blind interchange, followed by a discussion of MARTIF and associated research that provide a foundation for Blind MARTIF. The work continues with an overview of the current Blind MARTIF proposal, including pointers to supplementary materials on CD-ROM and the Internet.

To test this thesis, externally supplied data files are analyzed. Specifications of each data sample are submitted to the coder of the conversion routines. Finally, an analysis of the results of running the conversion routines is conducted, clearly showing that Blind MARTIF is an adequate approach to preservation of information.

ACKNOWLEDGEMENTS

This thesis would have been impossible without the support and encouragement of many people.

To my wife, Jill, and my children, Zachary, Wesley and Cierra, who not only helped me complete this process, but humbly and patiently endured many long nights, my loving thanks.

To John Watkins, Leah Blackburn, and Troy Cox, who were gracious enough to help me with this project on their own time.

To Daniel Hardman, whose thesis and continued work on the Blind MARTIF concept made my thesis possible, my sincere admiration.

To Ryan Corradini and Arle Lommel, who continually gave of their time and resources to help me complete this project, without which the thesis would have never been completed, my humble gratitude.

To Dr. Alan Melby, whose endless patience, ceaseless encouragement, innumerable suggestions, and endless enthusiasm for the ideas of the thesis, my unreserved appreciation. To Sue Ellen Wright, Klaus-Dirk Schmitz, Deryle Lonsdale, John Robertson, and Don W. Chapman, who offered important counsel and insight, my genuine appreciation. To the scholars and professionals who have developed MARTIF, my respect.

Table of Contents

1 Preliminaries to the experiment in blind terminology interchange

1.1 Introduction: Two options for terminology interchange

Translation tools are used extensively every day in the translation industry. Among these tools are those that are used to manage terminology by enabling translators to "remember" the terminology of past translations and translation standards. This can greatly increase the consistency and quality of translations. These tools use their own proprietary system for representing terminological data. Naturally, this can cause problems when two different data models (i.e., two particular conceptual data models within some approach to data modeling—see glossary) are used to represent terminology and there is a need for exchanging terminology between the formats.

There are two possible solutions to the problem of many existing translation tools, their many different formats for representing the terminological data, and the need for interoperability (see glossary):

1. The first is to somehow enforce a uniform internal data model for all terminological databases (termbases). Implementing the same internal design would greatly simplify the task of exporting and importing termbases. The problem is that there are already numerous internal data models in use in the translation industry, making it virtually impossible to enforce such a uniform internal data model. Who would have the authority to define and enforce such a data model? In the translation industry, there are many different and particular needs that are met by one data model or another. A flexible system such as MultiTerm™ allows a multitude of data models to be implemented, potentially a

different one for each user. To somehow require all termbases to have the same uniform internal data model would be met with great resistance in the industry.

2. The other possible solution for interoperability is a terminology interchange format that could serve as an import/export bridge between termbases with different data models. Using word-processing software, for example, it is possible to exchange documents between computers, as well as formats and platforms, using Rich Text Format (RTF – see glossary). No such format has been widely accepted in the translation industry for termbases up to this point. While RTF has its detractors, the concept of a blind interchange format is sound: it is the implementation of such a concept that is often bad. The same idea applies to MARTIF (MAchine-Readable Terminology Interchange Format – see glossary). The success of MARTIF as an interchange format is clearly dependent upon the ability of the user to implement it correctly.

It is assumed in this thesis that the second solution to the problem of interoperability, i.e., a terminology interchange format, is the only realistic solution. This work is intended for two audiences: (1) those who want to get a general understanding of terminology interchange and the particular experiment reported on here and (2) those who want an in-depth understanding of MARTIF. The second group will need familiarity with additional materials, including SGML (Standard Generalized Markup Language – see glossary) and XML (Extensible Markup Language – see glossary), a subset of SGML, and materials available on the CD-ROM provided with this thesis. Finally, the author will try to prove through evidence and analysis that blind terminology interchange (see glossary) is possible and practical through the utilization of Blind

MARTIF as an interchange format. Blind MARTIF may be submitted as a proposal to ISO for MARTIF Part 2 but is currently an informal project—a restriction on Negotiated MARTIF.

1.2 Tension between blindness and preservation (of information) in terminology

1.2.1 Motivation for blindness

As mentioned in the previous section, this work assumes that an interchange format will be used rather than attempting to impose a uniform internal data model on all translation tools. Within the interchange option there are two approaches.

One approach to terminology interchange is called "negotiated" interchange. In this approach, two partners use a common framework for interchange and negotiate details within the framework of the intermediate format to allow writing export and import routines that preserve as much information as possible.

Another approach is called "blind" interchange. With this approach, details are pre-defined so that export and import routines can be written without knowing who the other interchange partners will be (i.e. one cannot 'see' the interchange partners). The advantages of a blind interchange format quickly become apparent. Consider a student who has an assignment that must be submitted to the instructor in a file on a diskette. Additionally, suppose that the instructor requires the format of that file to be in RTF. What are the advantages to the instructor if everyone submits their documents in RTF? One advantage is that most word-processing programs today are able to export a document into RTF without having to worry about what type of word-processing program the instructor or any other user has on the other end. Secondly, RTF tries to preserve most of the formatting, thus minimizing loss of information. Thirdly, the instructor's software can then import the RTF

file, never necessarily knowing the format of the original file. Assuming that the negotiation of a document exchange format specifically for the course is out of the question, the alternative to blind RTF interchange would be to require all students to use the same version of the same word-processing software. This would be akin to using the same internal data model for all termbases.

It is well known that some have had difficulties exchanging documents using RTF. The difficulties that occur are often due to the fact that the RTF standard seems to continually change, making it difficult for software manufacturers to keep track of the current version. Although manufacturers may have difficulty fully implementing the current standard, the principle of using RTF is sound. In the translation industry, for example, RTF documents are often sent between individuals, e.g. project managers, translators, proofreaders, etc., when there is a chance that all parties will not have the same software for editing a document. Problems do occasionally surface, but generally this is a sound practice.

Compilers and users of terminology are also motivated to create systems that enable them to exchange information with a few or many other users. Despite this fact, there are no widely used machine-readable interchange formats that retain the structure and content of a termbase.

Nearly all terminology management systems are able to export data, most often in the form of a text file. On the other hand, almost none of them have the ability to view or import foreign data (i.e. from an external database), unless some form of negotiated interchange has been set up to facilitate this. What this means is that terminology management systems are very capable of sending others data, but those receiving the data are not necessarily able to read or import that data (Hardman, 1996:19).

Hardman (1996:20) describes how one solution to this problem would be to employ computer programmers to write conversion routines from each source data format to each of the target data formats. Although there is no question that such programming is possible, the amount of coding is not only excessive but also inefficient. Hardman explains what the implications of such a plan would be. If there were users from *n* different systems that each want to seamlessly import from and export to every other system, there would need to be a minimum of *n*(n-1)* different conversion routines (Figure 1) (Hardman, 1996:21). This equation shows how extensive the programming would be. If there were only ten (n=10) systems, there would be ninety programs; n=20, three hundred eighty programs; n=30, eight hundred seventy. Further, Hardman points out that this is a 'best case scenario'. The problem multiplies each time one of the formats changes, such as with version upgrades and new releases of software, resulting in changes to *2*(n-1)* conversion routines.

Figure 1: Conversion routines required with six file formats (Hardman, 1996:21)

A→B	A→C	A→D	A→E	A→F
B→A	B→C	B→D	B→E	B→F
C→A	C→B	C→D	C→E	C→F
D→A	D→B	D→C	D→E	D→F
E→A	E→B	E→C	E→D	E→F
F→A	F→B	F→C	F→D	F→E

The volume of programming required for such a strategy illustrates well why there is a need for a blind interchange format that can serve to facilitate the interchange of terminology. For example, if each terminology management system were only required to have conversion routines to export and import an accepted interchange format, the volume of programming would be only *2*n*. For n=10, only twenty conversion programs would have to be written as opposed to ninety. Additionally, only two programs would have to be updated,

i.e. the two conversion routines for that particular system, when one of the systems changes. If the preservation of information can be maximized, clearly a blind interchange format would be preferable to a negotiated interchange format.

1.2.2 Motivation for preservation of information

The goal of any terminology interchange format, whether negotiated or blind, is to minimize information loss. It is easy to imagine an interchange format that is entirely blind but where information loss is significant. For example, conversion routines could be written for a format that includes only three fields per entry: entry subject field, term 1 (source), and term 2 (target). However, we must consider the consequences of the information loss involved in such a simple blind interchange format. All personal comments, all administrative data about the quality of the entry, all definitions and explanations, all contextual examples of the entry would be lost with such a paltry implementation of a blind interchange format. Such information is important—particularly to system managers—and cannot afford to be discarded (See Appendix G, which contains arguments for doing careful terminology management). A blind interchange format that is successful, i.e. adopted by industry, must necessarily allow for as much information preservation as possible. Any additional information in a source entry beyond the three fields of the simple exchange format just mentioned could just be attached to the end of the entry in a field called "other". However, passing data along without indication of the data type does not constitute preservation, since translation tools would not be able to use the data automatically.

1.2.3 Blindness vs. Preservation

It is clear that the above three-field format would succeed in allowing blind interchange, but at the price of excessive information loss. On the other hand, a format that simply specified units of text nested in units of text with a type attribute on each unit as well as user-defined types would practically eliminate information loss in negotiated interchange (that is, the interchange would be "complete"). This, though, would make blind interchange impossible, as the structure of such an interchange format would be so fluid that a blind import routine could not be written. Somewhere between these two extremes of total blindness with considerable information loss (i.e., blindness without preservation) and negotiation with total preservation of information (i.e., preservation without blindness), there is hopefully a format that supports blind interchange without excessive information loss.

The need for either blindness or negotiation in terminology interchange can be viewed as an application of the general linguistic principle of inverse proportionality (Robertson, personal communication). Specifically, the degree of constraint on input (i.e., the form the import data can take) is inversely proportional to effort to decode (i.e., the ability of the machine to "understand" the import data). For example, consider the effort needed for a computer to decode free-form handwriting. Technology has begun to afford humans the ability to input data that a machine can recognize and manipulate via handwriting, such as with the use of a personal digital assistant (PDA). We see, however, that the restrictions placed on PDA users are significant, to the point of instructing the users about how they are to input data, i.e. writing the alphabet. This requires from the "decoder"

(PDA, in this example) less effort to "understand" the input. If the PDA did not constrain the manner for inputting the data, the PDA would require much more effort to "understand" the input, i.e., the software to handle the writing would be much more complex.

Likewise, negotiated and blind interchange are related in that they impose two methods for constraining the input, which in this case is the interchange file presented to the import routine. The more defined the constraints, whether by negotiation between two partners or by acceptance of the constraints in the specification of a blind format, the easier it is to write an accurate decoder (that is, an import routine that preserves the information in the interchange file). The looser the constraints (fewer constraints) the harder it is to decode. With no constraints, the effort to decode the input becomes infinite, like dividing by zero (Robertson, personal communication).

The battle is to define how information can be preserved without losing data. The more constrained a system is (i.e., the more defined its import format is) the more prone it is to information loss. For instance, many high-visibility Web sites offer advertising banners to be purchased and placed on the site itself. Most sites require that the format of such banners meet certain requirements, such as its physical dimensions in pixels and the amount of memory in bytes the banner may have. Generally, the problem does not lie with meeting the pixel requirement, rather meeting the memory requirement. To meet the memory requirement, the banner owner may decide to display a smaller range of possible colors in the banner, e.g., 16 vs. 256, resulting in fewer bytes. Naturally, the banner will lose some detail, that is, with more constraints, less preservation. However, the question is how to decide what amount of information loss is acceptable if at all.

In section 1.3, I examine what progress in the preservation of information had been made in negotiated terminology interchange, specifically that of Negotiated MARTIF. In section 1.4, I explore a proposed blind terminology interchange format that is based on Negotiated MARTIF and the suggestions made concerning the Hardman thesis. In section 1.5, a methodology to test the adequacy of this proposed blind interchange format is explained. Finally, in section 2, the test is conducted and reported on.

1.3 Foundation for a blind terminology interchange format

1.3.1 Need for a data model

Most terminologists would agree that there is a need for a data model underlying every termbase. Holmes-Higgins and Ahmad (1996: 215-224) point out the importance of providing explicit data models for all types of terminological data. A data model is a description of (a) types of data elements (known as data categories in terminology work), i.e., what types of information are allowed in a termbase, and (b) their relationships. These two features are essential to the basic definition of a data model. Suppose that a data model had no method for indicating relationships between data elements in a termbase. The resulting termbase would not be a termbase at all: it would simply be a list. For example, if a term is followed by a date, but there is no mechanism for relating the date to the term versus the whole entry, then the objective of explicitly representing information in the termbase has been defeated—without human intelligence there is no way to even postulate what the date applies to. Further, suppose that a data model had no method of describing the data elements of a termbase. Again, the result would simply be text and of no particular use in terminology interchange. Naturally, most if not all termbases have a data model. Consequently, an interchange format must also have a data model.

1.3.2 ISO FDIS 12620

System managers use a wide variety of data categories for storing and retrieving terminological data. ISO FDIS 12620 (Computer Applications in Terminology – Data Categories) inventories the data categories used by system managers. A benefit of having 12620 is that it can be used as a reference by current and future system managers that encourages them to use common data categories within their termbases in the likelihood that eventually they will have the need to import and/or export terminological data.

The remainder of this section is referenced and is a summary of the information found in the CLS Framework (Concept-oriented with Links and Shared References – see glossary). The data categories of ISO FDIS 12620 are divided into ten sections that are grouped into four classes: term, term-related information, descriptive information, and administrative information. The intention of 12620 is not to show how these data categories are structured within a termbase. The implementation of the data categories is left to the individual system manager. In a particular data model several 12620 data categories may become permissible instances of a single implemented data category, such as the various types of term. In the rest of this section I will discuss how this standard may be represented in SGML and in a relational database respectively. At this point, however, I will simply discuss what data categories are included in 12620 and how they are divided into classes.

Term is the first class of data categories and only consists of the data category term. This data category is part of a concept entry (an abstract notion designated by one or more terms). Each term can be modified and further categorized using the items listed in the second class of data categories, term-related information.

Term-related information, the second class of data categories, is attached to a term

and indicates specific properties of the term. Among others, information such as part of

speech (e.g., verb, noun, etc.), register, etymology, and administrative status are attached to

the term. Term-related information also deals with issues of equivalence, such as transfer

comments, which are also attached to the term.

Descriptive information is the third class of data categories. These data categories

describe a concept. They include subject field and position in a classification system,

descriptive information (such as definitions), contextual examples, explanations, and

relations between pairs of concepts. For example, a lens is part of a camera, so a *lens*

ConceptEntry (see glossary) could have the data category *partitive relation* attached to it.

The content for the instance of the partitive relation data category would then include *camera*

and a link to the *ConceptEntry* for *camera*. Alternatively, there are data categories that relate

a *ConceptEntry* to its position in a concept system. In an external concept system separate

from the termbase for cameras, the *ConceptEntry* for *lens* could include a link to a node in a

concept system which would indicate that a certain node in that concept system is the whole

for which lens is a part (Melby, 1998).

The last section of data categories within the descriptive information consists of the

data category *note*. It is recommended that more specific data categories be used to hold

descriptive information, but that is not always possible, in which case *note* can be used.

Administrative information is the final class of data categories. This information may

indicate such things as the documentary language used to characterize data or the content of

documents to permit their storage and retrieval (ISO FDIS 12620: 30). Additionally,

administrative information may include data categories that relate to a thesaurus, which can

facilitate the retrieval of documents by attaching thesaurus items to documents and indexing

on those items (Melby, 1998). Naturally, administrative information will include data

categories that indicate when things happened, when documents were updated, who created

them, who updated them, etc.

The purpose of ISO 12620 is not to actually specify what data categories should or

should not be used in a particular termbase. The purpose is to detail what types of data

categories are typically found in local database systems, and to facilitate terminology

interchange between systems. This can be accomplished by mapping local system data

categories to the universal data categories specified in this standard (see Figure 2). Further,

the negotiated and blind forms of an interchange format that are based on this standard will

be discussed in sections 1.3.3 (the CLS Framework), 1.3.4 (Reltef™), and 1.3.5 (Negotiated

MARTIF).

Figure 2: How an interchange format can link different data formats
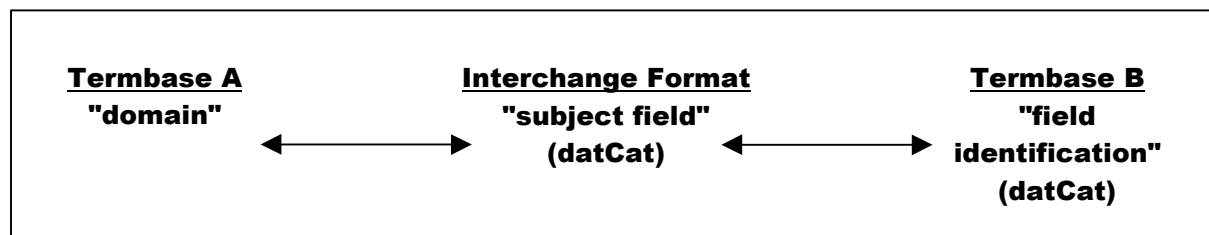
1.3.3 The CLS Framework

Blind MARTIF, to be defined in section 1.4, is based on a joint effort between the

Brigham Young University Translation Research Group (BYU TRG) and the Kent State

University Institute for Applied Linguistics (KSU IAL). This collaboration has resulted in the

development of the CLS Framework (Concept-oriented with Links and Shared references).

Whereas ISO FDIS 12620 deals only with an inventory of data categories found in typical

termbases and does not specify the relationships among items in an entry, the CLS

Framework deals with both data categories (taken from ISO 12620) and with the structure

and relationship of the data categories.

As Melby explains (1998), a termbase compatible with the CLS Framework consists

of (a) global information about the termbase, (b) a set of concept entries (called

*ConceptEntry*s in this document), and (c) a set of references (here called *SharedRef*s – see

glossary) that can be shared among multiple *ConceptEntry*s or parts of *ConceptEntry*s.

Terminological data may be represented in at least two ways. A termbase may be in

the form of an SGML file or in the form of a relational database. The CLS Framework can

be used to create a "family" of compatible data models (Melby, 1998) using the data

categories defined in ISO FDIS 12620. It is already used in Negotiated MARTIF (see

section 1.3.5), which is an SGML approach to terminology interchange, and in Reltef™ (see

section 1.3.4), which is a relational approach to terminology management. Since both

Negotiated MARTIF and Reltef™ are consistent with ISO FDIS 12620 and with the CLS

Framework, it is possible to write automatic bi-directional conversion routines between them

with maximum preservation of information.

As mentioned in section 1.3.2, ISO FDIS 12620 describes four major classes of data

categories: *terms*, *term-related data categories*, *descriptive information*, and *administrative

information*. The CLS Framework describes how these classes of data categories can be

structured to show their relationships in entries and used to link global information (i.e.,

information that pertains to the entire termbase) and shared information (i.e., information that

is shared among certain *ConceptEntry*s).

Global information might include the name of the database, information about the owner(s) and updater(s) of the database, dates and version numbers (Melby, 1998). As Melby (1998) points out, some or all of this information may be placed outside of the termbase itself, although this is not encouraged, as it may hinder terminology interchange.

A *ConceptEntry* ideally consists of information about a single concept and lists the language-specific terms associated with that concept. The bulk of the information in a termbase is in a set of ConceptEntrys that, in the context of the CLS Framework, constitutes the body of the termbase.

Ideally, the terms in a *ConceptEntry* will represent the same concept within a particular subject field. Terminologists will agree, however, that it may often be difficult to find perfect equivalents between languages, especially if the terms are cognates (Hardman, personal communication). Transfer comments may be needed to indicate the degree of equivalency between the terms of a *ConceptEntry*. If the degree of equivalency falls below some threshold set by the terminologist, it may be appropriate to define two separate *ConceptEntry*s. As Melby (1998) points out, the ideal *ConceptEntry* has the following properties: the entry is based on a single, well-defined concept, and all the terms in the entry are equivalent to each other. The goal may be to have one term in each language for each *ConceptEntry*, but if this is not possible, e.g., synonyms exist, this should be documented to reflect the degree of equivalency.

In addition to global information and *ConceptEntry*s discussed above, *ConceptEntry*s may include links to other items in the *ConceptEntry*, other *ConceptEntry*s, or to *SharedRef*s. For example, the same terminologist may have modified a number of ConceptEntrys. It would be efficient to have a link to the biographical data (biodata) of the responsible party in

the *SharedRef*s (centralized location), rather than have the same information present in each *ConceptEntry*. When using this method, making an update to the termbase is much easier, particularly since only one element needs to be updated.

The use of *SharedRef*s follows a sound principle of database management: redundant information should be avoided. Melby (1998) explains several reasons for this practice:

1. Economy of storage: As mentioned above, several *ConceptEntry*s can point to the same detailed information that pertains to them all.

2. Consistency: Updating of information is much more efficient, if only one element must be updated. Otherwise, if several *ConceptEntry*s have the same detailed information attached to each; care must be taken to find every occurrence of the information in order to make the change in all locations.

3. Flexibility: Each *ConceptEntry* can point to the *SharedRef* and add the specific information that is distinctive to that reference, for example, a page number in a book (Melby, 1998).

Links to *SharedRef*s within a termbase can thus take two forms to connect data items:
1. Simple pointers that link data entries, data segments, or individual data elements to the *SharedRef*s.

2. More complex links that contain additional information in the form of variant text, such as a page number indicator or other information on which part of a reference is relevant to the item in question.

It is important to mention that *SharedRef*s can be used to link relationships between *ConceptEntry*s themselves. In order for this to be possible, each *ConceptEntry* must have an identifier, used as a pointer reference that is unique throughout the termbase.

These three components (global information, *ConceptEntry*s, and *SharedRef*s) along with unique item identifiers and a link mechanism, constitute the minimal features of an effective termbase.

1.3.4 Reltef*ä*

As mentioned above, Reltef™ is an implementation of the CLS Framework. It includes an entity relationship diagram and can be implemented in any medium range to high-end relational database management system.   For more details, please see http://www.ttt.org/clsframe/reltef/.

1.3.5 ISO FDIS 12200 - Negotiated MARTIF

ISO FDIS 12200, also called Negotiated MARTIF, is another very flexible and promising data model compatible with the CLS Framework. Negotiated MARTIF has resulted from many years of international cooperation, beginning with the Text Encoding Initiative (TEI) and the Localization Industry Standards Association (LISA). As Hardman (1996:24) points out, the intended purpose for Negotiated MARTIF is to facilitate the negotiated interchange of terminology among various systems, applications and platforms.

Negotiated MARTIF is based on ISO 8879, Standard Generalized Markup Language (SGML) (ISO 8879: 1986) and has been designed to be as open and flexible as possible to accommodate as many elements and forms of termbases as possible. The data category (i.e. element) structure of Negotiated MARTIF is partly defined in a document called a DTD (Document Type Definition - see glossary). A DTD defines the data items that are permissible in a document of a given type and how these items relate to each other.

Clearly, it is important to understand how important a DTD is in maintaining the structure and consistency of any SGML-based document. An example of the use of a DTD in the every-day life of many people, perhaps without even knowing about it, is the way they view Web pages on the Internet. Every time someone accesses the World Wide Web for any purpose, there is an HTML (Hyper Text Markup Language) DTD that explains to the computer, more specifically to the browser software (e.g., Netscape Navigator®, Microsoft Internet Explorer®, and others) how to represent Web pages for viewing by a human on the computer screen. The internal structure of a Web page is based upon an HTML DTD that designates what elements can be part of a typical HTML document and how these elements can be structured, i.e., attached to each other. For instance, a <p> (paragraph) tag may not be required to have a close tag, i.e., </p>. Further, a <p> tag may be embedded in an <ol> (ordered list) tag. The HTML DTD establishes the usage of this and all other HTML tags. Similarly, the MARTIF DTD lists the tags allowed in a MARTIF document and their relationships.

1.4 Defining Blind MARTIF

1.4.1 Restricting Negotiated MARTIF to create Blind MARTIF

As mentioned in section 1.3.5, Negotiated MARTIF can facilitate the negotiated interchange of terminology between partners who understand each other's data model. Since Negotiated MARTIF does not rigidly specify the allowable values for data categories in advance, it leaves it to the negotiation step. That dramatically reduces the possibility of blind interchange, in which a number of partners want to be able to export and import terminological data without "seeing" the interchange partner (Melby, 1998).

The definition of blind interchange implies that there is total independence between interchange partners. Data is represented in such a manner that all possible sources of unpredictability are eliminated, enabling fully automatic conversions without regard for the source of the terminology. Current technology does not allow for the intuitive interpretation of data as a human might do when presented with an ambiguity of meaning (Alder, 1997; Hardman, 1996); therefore, blind interchange must presuppose a completely unambiguous and predictable format.

Hardman (1996) clearly explains various reasons why the flexibility of Negotiated MARTIF may not always be conducive to blind interchange:

1. Negotiated MARTIF provides no blind mechanism for encoding custom character sets and non-alphabetic languages.

2. Often two pieces of information are needed to make a data element meaningful, whereas Negotiated MARTIF may not require the full construction.

3. Negotiated MARTIF defines how data is contained, but often may not provide information about the data itself. For example, tags such as <termNote type='grammaticalGender'> in Negotiated MARTIF offer no clues as to how the gender identification must be structured, e.g., *m*, *mas*, *masc*, etc. If the target termbase allows only certain values in its grammatical gender, then blind import routines cannot map into those values unless the possible values in MARTIF are pre-defined. Many Negotiated MARTIF elements allow the interchange of system-specific information, but describe nothing about how that information must be structured within the content of the element.

4. Negotiated MARTIF is structured such that identical pieces of information can be encoded in two or more different ways (ISO FDIS 12200.2, Annex D), i.e. it allows data modeling variance.

5. Negotiated MARTIF may point to information that is external to the document itself. If such a MARTIF document is converted into another format and either the external data is not transferred with the document or the new format does not support external data, then information will become invalid.

6. Negotiated MARTIF made blind interchange difficult because data may be placed in different orders, sometimes representing intricate relationships that may be intuitive to a human, but are difficult or even impossible for conversion routines to infer. This problem, which was in the 1995 draft version of 12200, was resolved with the 1997 version of 12200.

Given the fact that Negotiated MARTIF displays undesirable flexibility, the following sections specifically address how Negotiated MARTIF has been sufficiently constrained to facilitate our objective of blind interchange. It is important for reasons of historical continuity and stability to build a blind interchange format on the existing standard of a negotiated format, if possible, rather than beginning with an incompatible format. Appendix A of this paper contains the detailed specification for blind interchange, i.e., Blind MARTIF, including the DTD described in section 1.4.4.

1.4.2 Coming in line with XML & Unicode

Recently, there has been much talk about a new subset of SGML called XML (Extensible Markup Language), which has been developed under the auspices of the World

Wide Web Consortium, with such major players as Microsoft and Sun (World Wide Consortium, 1998a). XML describes a class of data objects called XML documents and partially describes the behavior of the computer programs that process them. XML is an application profile or restricted (subset) form of SGML (Microsoft, 1998). XML is intended to be usable over the Internet and to support a wide variety of applications. It is intended that XML documents should not only be easy to write, but should also be human-legible. The advantages of making Blind MARTIF XML-compliant are that:

- first, XML has one standard method for representing character sets, and

- second, it will facilitate the automatic and reliable interchange of data (Mace, Flohr, Dobson, Graham, 1998), specifically with the writing of conversion routines.

In order to come in line with XML, some elements of the Negotiated MARTIF DTD would be required to have end tags, where they currently are optional. For example, elements such as *p*, *titleStmt*, and *publicationStmt* would now be required to have end tags, in one form or another. Also, empty tags or tags that syntactically do not have end tags in XML have a slash at the end: <ptr/> (end pointer, i.e. pointing to other, related information within the MARTIF document). This makes it clear to software that there is no need to look for an end tag, even without consulting the DTD. Together with the prohibition on implicit end tags on elements that do have content, these restrictions make XML much easier to process than unrestricted SGML. This affects only elements in the header, since Negotiated MARTIF contributors originally required end tags on all the elements used in the body. For example, in Blind MARTIF a header might be:

```
<publicationStmt><p>not published separately</p></publicationStmt>
<sourceDesc><p>from ISO DIS 12200, body: Example 10</p></sourceDesc>
```

Note that each start tag in the header example above has an end tag, where before in Negotiated MARTIF it was not required.

Concerning international character sets, a PC with a Windows95/98 operating system (OS) utilizes a different character set than does an Apple Macintosh. In fact, Windows95/98 and Macintosh operating systems can simultaneously display several character sets on the screen. Further yet, mainframe computers will use still a different character set. The only thing they may have in common is a subset of the ISO 646 standard, which contains only English characters (Melby, 1991). This can obviously create difficulties in the ability to exchange information between platforms, something that is obvious to those who work with two or more platforms.

The Unicode Worldwide Character Standard is a 16-bit character coding system designed to support the interchange, processing, and display of the written texts of the diverse languages of the modern world.   In its current version, the Unicode standard contains 38,885 distinct coded characters derived from the supported scripts, such as Hangul, Arabic, and Thai among others (Unicode, 1998).

Negotiated MARTIF has no standard method for representing languages that employ non-English characters. Therefore, it is impossible for a conversion routine to automatically determine what encoding standard is used in a Negotiated MARTIF document. Blind MARTIF, in compatibility with XML, requires the Unicode codeset, and goes one step further by specifying two and only two Unicode encoding schemes. The first of these is UCS-2 (two bytes per code point). The second is tentatively referred to as UHA (Unicode Hex ASCII), as it is simple ASCII with hex character entities. In this encoding, Unicode codepoints under 128 are represented as ASCII and codepoints of 128 and above are

represented as hex character references using sequences of eight ASCII characters. For example, & (ampersand) signifies that an *entity* follows, # (number sign) indicates that a number follows, and *x* indicates that the number is hexadecimal (as opposed to decimal). Finally, the *;* (semi-colon) signifies the end of representation. Thus, the Unicode codepoint "3FC hex" would be represented as "ϼ". Formally, UHA is a sequence of ASCII characters, some of which are part of hex character entities. In UHA, mnemonic character entities in addition to the standard XML entities, such as &lt, &amp, and &gt, are also allowed, but only if their value is a hex character reference.

1.4.3 Blind MARTIF Data Categories

1.4.3.1 Maxi-set

The maxi-set of data categories is the total set of data categories (from ISO FDIS 12620) allowed in Blind MARTIF. In this section, a brief description of each class of data categories, with indicative examples of how data categories are represented in Blind MARTIF, and direction where further detailed information can be found about the individual data categories is given. In general, information about the data categories can be found in either ISO FDIS 12620 or in the CLS Framework at http://www.ttt.org/clsframe/. It should be noted that the content of some data categories may be unrestricted, while others may be restricted with a list of permissible instances. Additionally, extra data categories may arise or be required. Two such data categories are *maxi* and *custom* and are explained in section 1.4.3.3.

Term

As mentioned, there is only one data category in this class: *term*. It is expressed in Blind MARTIF in the following way:

<term>ink jet printer</term>

Term-related information

The following three data categories are indicative of the variety of categories available in this class. A full listing is available at the Internet site mentioned above.

Part of Speech:
<termNote type='partOfSpeech'> noun </termNote>

Etymology:
(together with the term *telecom*)
<termNote type='etymology'> telephone + communications </termNote>

Reliability Code:
(where reliability is expressed by a number between 1 and 10)
<termNote type='reliabilityCode'> 6 </termNote>

Descriptive information

Subject Field:
(here, the colons show the hierarchy of subject fields: "diseases" is a sub-field of "medicine", and "cancer" is a sub-field of "diseases")
<ref type='subjectField' target='medicine'> medicine:diseases:cancer </ref>

Definition:
(together with the term *telecommunications* in a <descripGrp>)
<descrip type='definition'> long-distance communication by electronic means </descrip>

Partitive Relation:
(together with the term *lens* in a <descripGrp>)
<ref type='partitiveRelation' target='Idcamera'> camera </ref>

Administrative information

Update Date:
  <date type='modification'> 1998-02-14 </date>

Originator:
(Where KDS is the ID of a bibliographic reference of type responsibleParty, i.e. Klaus-Dirk Schmitz)
  <ptr type='originator' target='KDS'/>

1.4.3.2 Constraints

  One of the primary strengths of Negotiated MARTIF is its flexibility. However, as has been explained, this becomes a liability in blind interchange, as it creates the problem of unpredictability in representation. To resolve this, three main constraints have been applied to Negotiated MARTIF to facilitate the blind interchange of terminology.

  Content of Elements

  Recall the example of *grammaticalGender*:

  <termNote type='grammaticalGender'>…</termNote>

In this example and others, the content of the element is unrestricted by the DTD, and in fact cannot be restricted by the DTD.

  If blind interchange is to be possible, however, elements such as these must have concrete definitions. A routine to import from Blind MARTIF must be able to make the assumption that the content of a *grammaticalGender* element will be "masculine," "feminine," "neuter," or "other," with no other possibilities. If, for example, a Blind MARTIF document contains the following structure:

  <termNote type='grammaticalGender'>masc.</termNote>

24

an import routine would not recognize this as a valid gender, and the information would be lost (see Figure 3), since the routine would not know how to pass the information on. It may be possible for one to program the conversion routine to recognize a number of forms of the content, but this would complicate the writing of all import and export routines, particularly if one takes into account that this is just one of many possible instances where the content of a given element should require a concrete definition.

Figure 3: The necessary mapping of the content of certain data categories

SOURCE	**Blind MARTIF**	**TARGET**
"masc"	"masculine"	"m"
(gender)	(gender)	(gender)
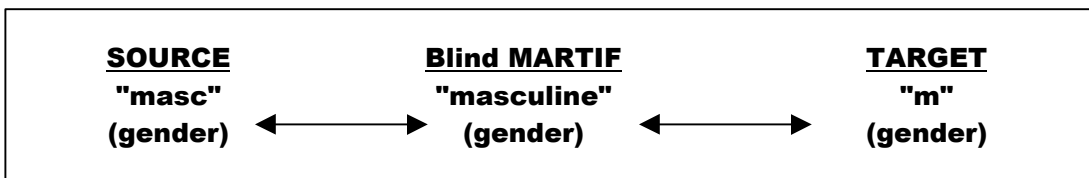
Level

 Additionally, some elements must be restricted to certain positions within a document's structure. For instance, it makes little sense to maintain that a homograph should occur at the *conceptEntry*-level. The *homograph* element should only occur in conjunction with a specific term, i.e. at *term*-level.

Target

 Finally, the targets of cross-reference elements (ptr and ref) must also be restricted. For example, a bibliographic source reference related to a specific term should be required to point to a document reference (*refObject*) in the back matter, and an element indicating an entry's creator should only be allowed to point to an element describing that person.

 Some of these constraints cannot be asserted in the Negotiated MARTIF DTD itself; however, permitting these elements to go unrestricted would allow unacceptable, ill-formed structures to be created and passed off as "valid" data. Some constraints could be expressed

in a greatly modified Negotiated MARTIF DTD, but this would hide common features within

similar structures. The SGML/XML community accepts that the work of validating a format

can be shared between the SGML DTD and other formal specifications (Michael Sperberg-

McQueen, personal communication; World Wide Web Consortium, 1998b; Microsoft, 1998b),

such as a DCS (data category subset - see glossary) file.

The following example is an extract from the DCS file. Specifically, the *repeatability*

element defines whether a Blind MARTIF document may have multiple instances of the

respective data category at the same point in the document. The target element applies only

to the data category with a target attribute and defines where it can point. Blind MARTIF

represents these constraints in the following XML element:

```
<refObject type='datCat'>
   <item type=GI>termNote</item>
   <item type=typeVal>termType</item>
   <item type=content>"main entry term", "synonym", "international scientific term",
"common name", "internationalism", "full form", "abbreviated form", "variant",
"transliterated form", "transcribed form", "symbol", "formula", "equation", "logical
expression", "sku", "part number", "phraseological unit", "standard text"</item>
   <item type=levels>termGrp, termNoteGrp</item>
   <item type=repeatability>*</item>
   <item type=targets>*none*</item>
   <item type=positionNumber>A.2.1.x</item>
</refObject>
```

A "*refObject*" is a specific type of MARTIF element, used for storing data general to an

entire document, such as terminologists' biodata bibliographic references, and in this case,

data category (datCat) constraints. Together the GI and typeVal items make up the unique

data category name, and the other three items prescribe its behavior.

The complete set of these datCat definition *refObject*s is collected in a *refObjectList*,

which can then be referenced either explicitly or implicitly by a Blind MARTIF document.

In this way, an external program may be used to verify the "validity" of a document. This is explained in Section 1.4.5.

External references may be included in a Blind MARTIF document by using the *xref* tag with an optional note as the content of a *refObject*. The value of target on *xref* is a namespace ID (URL or FPI), with the content being a description of where to look in the targeted document.

1.4.3.3 User subsets

Blind MARTIF includes two kinds of restrictions:

- a restriction on 12620 data categories (the maxi-set is a subset of 12620 and user-group subsets are subsets of the maxi-set); and

- a restriction on the DTD in a way that assures that any Blind MARTIF file would parse with the Negotiated MARTIF DTD (assuming an XML-friendly SGML declaration and a parser that accepts hex character references).

Blind MARTIF is a restriction but also an augmentation of Negotiated MARTIF. It includes a general-purpose Blind MARTIF validator that reads the <encodingDesc> element to obtain the name of the datCat subset being used and then obtains that subset definition in the form of a piece of XML text and interprets it as a logical restriction on Blind MARTIF, including a restriction on the datCats allowed and the DTD options allowed by restricting placement of datCats to certain levels in the DTD. So "Blind MARTIF Maxi" would be the full name of what we call just "Blind MARTIF," which is MARTIF restricted to the Blind MARTIF DTD and the maxi datCat set, along with utilities that apply to various further restrictions of Blind MARTIF. A restriction of the maxi datCat set, along with any relevant

software and other gateway formats, are called "user subsets" of Blind MARTIF (Melby, personal communication).

As mentioned in section 1.4.3.1, it is conceivable that members within the same user subset group may have certain, extra data categories that are not included in their particular data category subset. There are two possible situations where this might occur. First, the source termbase might have a data category that is not included in the user subset, but is part of the maximal set of data categories (see Appendix A.3 - Maxi-Set). In this case, the Blind MARTIF representation would be:

<ref type='maxi' target='X'>permissible instance</ref>, where X is the *refID* (see glossary) of a <refObject> that includes the data category from the *maxi-set*.

Second, the target termbase might have a data category that is neither part of the user subset nor the maxi-set. The Blind MARTIF representation would be:

<ref type='custom' target='X'>…</ref>, where X is the *refID* of a <refObject> that includes the custom data category from the source termbase.

Naturally, it is strongly encouraged that a maxi-set data category is used rather than a custom data category. This will at a minimum allow the human to look up the data category in the maxi-set and find a definition. A custom data category does not allow this. Although none have been selected as of yet, user subsets will be developed in the future. One such group called MARTIF Lite is currently being developed. It appears that the Localization Industry Standards Association (LISA) via the OSCAR special interest group (Open Standards for Container/Content Allowing Re-use) will be among the first group to request a user subset.

Even though the use of the entire data category maxi-set will allow for blind interchange of terminology, allowing for user subset groups to be formed is also

advantageous. Defining exactly what data categories are relevant to that particular user

subset group enhances the ability of Blind MARTIF to eliminate data categories that are

deemed unnecessary and perhaps ambiguous to a particular user. By doing so, programmers

would be able to more efficiently write conversion routines between members of the same

user subset. For example, if a user subset group has only twelve data categories rather than

the entire maxi-set, conversion routines would only have to account for those twelve. In

addition, a user subset may allow further control over the content of certain data categories,

such as *subjectField*, which cannot be defined for all users (Melby, personal

communication).

1.4.4 Blind MARTIF Document Type Definition description

As mentioned in the previous section, it is important to make any Blind MARTIF

document compatible with the Negotiated MARTIF DTD. A conscious effort has been made

to restrict the Blind MARTIF DTD enough to allow blind interchange, while allowing Blind

MARTIF documents to parse under Negotiated MARTIF.

Where type values of various Negotiated MARTIF elements are free and

unrestrained, Blind MARTIF often specifies these values, allowing for automatic machine

processing. This is accomplished in a DCE file, which is a logical part of the DTD (see

Appendix A).

Several cosmetic changes were made to the Negotiated MARTIF DTD to increase

readability and clarity in the Blind MARTIF DTD. These changes included alphabetizing

and grouping all element and attribute declarations, and introducing the *noteLinkInfo* entity

to simplify the element attribute definition. Further, as the Negotiated MARTIF DTD was

composed of three files, the Blind MARTIF DTD combined these files, increasing clarity of structure.

Blind MARTIF requires the use of the *langSet* (i.e. language set – see glossary) entity for explicit grouping of *ntig*s (i.e. nested term information group – see glossary) in the same language, where Negotiated MARTIF permits a *ntig* to stand on its own in a *termEntry*. Additionally, the *lang* attribute is now allowed on only three elements in the body of a Blind MARTIF document besides elements containing text (such as notes), and is required in all instances: MARTIF (for defining a document's meta-language), langSet (for grouping *ntig*s of the same language), and foreign (for marking foreign text in a note or descrip).

The MARTIF element of a Blind MARTIF document requires a version name, such as PART2V1 for "Part 2 Version 1." This avoids the problem that RTF faces (see section 1.2.1), namely that the version of RTF for two files may not be the same yet not distinguished from each other.

The remaining changes from the Negotiated MARTIF DTD are concerned with restrictions to and eliminations of MARTIF elements and attributes. To avoid data modeling variance, many redundant data representation options are eliminated. For example, the *tig* element (i.e. term information group) is eliminated in favor of *ntig*. This facilitates the grouping of related information (e.g. a definition and a definition source), while eliminating the need for adjacency. Additionally, the encodingDesc that is optional in Negotiated MARTIF is required in Blind MARTIF. The *lang* element is now a *refid* (see glossary) and must be defined in the Blind MARTIF document, where it is *CDATA* in the Negotiated MARTIF DTD (see Appendix A). This provides a list of the languages used in a given document.

1.4.5 Blind MARTIF Validation Overview

The validation of Blind MARTIF documents is described in detail in Appendix D of this thesis and on the Translation, Theory and Technology web site ("The 1998 Blind Interchange Experiment" - http://www.ttt.org/1998experiment/). In this section I discuss specifically what features of a Blind MARTIF document are checked for validity. If these criteria are met by a Blind MARTIF document, it is considered valid.

As described in section 1.4.4, steps have been taken to make Blind MARTIF compatible with XML. As software applications more widely adopt the XML standard, well-formedness checks of XML documents would be built-in. In order to be valid, a Blind MARTIF document must not only be well formed, but must be parsed according to the Blind MARTIF DTD. Further, the MARTIF element must have a registered type value in order to be valid. Currently, version 1 of Blind MARTIF is available. A Blind MARTIF document must also have a registered datCat set (i.e. the DCS file – see glossary), and the document must conform to that datCat set. It is planned to eventually establish a Blind MARTIF consortium, where users and user subset groups would be able to register datCat sets to further encourage interoperability.

1.5 Experiment Design

1.5.1 Methodology for developing conversion routines

In section 1.4, I described the Blind MARTIF data categories, specifically the Blind MARTIF DTD and the validation of Blind MARTIF documents. Section 1.5 proposes a methodology for determining the validity of the proposed Blind MARTIF format. First, I describe the writing of the conversion routines (section 1.5.2) from sample termbase formats into Blind MARTIF (SOURCE to Blind MARTIF) and Blind MARTIF back into all sample

termbase formats (Blind MARTIF to TARGET). Admittedly, the design of the experiment

is less than ideal since the same person pseudo-coded both conversions, which were then

used to write the conversion routines. A team of two would have been even better, but not

feasible due to the lack of manpower. Second, I discuss the methodology for determining

whether the conversion routines are successful at preserving information, while minimizing

loss (section 1.5.3).

1.5.1.1 Analyze data samples

The method for writing the export routines is straightforward and practical. An

analysis of the sample termbases is conducted to discover what data categories are being used

and how they might be mapped to Blind MARTIF. After the mapping has been completed, it

is good practice to pseudo-code the conversion of the sample data model, which in turn

facilitates the writing of the actual conversion routines.

1.5.1.2 Valid export routines

An export routine is considered valid if it can successfully map the original source

term entries into Blind MARTIF, while preserving as much information as possible.

Although there is often ambiguous content in the data categories of the sample systems,

Blind MARTIF must be able to account for as much of the information as possible. Each

Blind MARTIF file converted from the sample data is checked (1) for well-formedness and

(2) validity as has been defined in section 1.4.5.

1.5.1.3 Freeze export routines

Once the export routines have been written and tested with a portion of the externally-

supplied sample test data, each routine is frozen to ensure that there is no updating or

changing of the routines later to facilitate blind interchange and to avoid any appearance of negotiation.

1.5.1.4 Valid import routines

After the export routines have been written, tested and frozen, the import routines are written following the same procedures as were used to write the export routines. Each routine is designed to import any Blind MARTIF file into the sample database format, without 'knowing' where the sample came from. This, again, ensures blindness. Secondly, all Blind MARTIF files, regardless of their origins, will then be imported into each sample database format, to validate the blindness of the conversion, totaling sixteen new files, four in each of the supplied data models.

1.5.1.5 Freeze import routines

Again, to avoid any appearance of negotiation, the import routines are frozen after being written. It may be tempting to 'tweak' the programs later, if it is discovered that there is some unacceptable information loss or there is a belief that there might be a more efficient method to facilitate blind interchange. This temptation was resisted since this is simply an experiment to test the concept of blind interchange. Valuable information is bound to come forth after the test is complete about how blind interchange may be better facilitated. There is no guarantee of the success of this experiment. It is the premise of this work, however, that blind interchange is possible and advantageous using Blind MARTIF.

1.5.2 Applying the conversion routines

Roger Bennett, from the European Commission Translation Service, attended the ISO TC 37 meetings in August of 1997 and volunteered to supply sample termbase data to test

various proposals for a blind interchange format that may become a future ISO work item. Mr. Bennett has supplied Brigham Young University's Translation Research Group with sample test data extracted from four sources: GlobeDisk, Eurodicautom, and two different MultiTerm data models.

1.5.3 Criteria for evaluating the success of the experiment

1.5.3.1 Is the interchange truly blind?

As mentioned in section 1.5.1.2, the export routines are considered valid if they can produce valid Blind MARTIF. For the interchange to be truly blind, the import routines (i.e. Blind MARTIF to target database format) must necessarily properly convert Blind MARTIF documents converted from any source database format, i.e. the source cannot matter. One way of testing this is through parsing the Blind MARTIF files with a standard XML parser. This indicates whether the Blind MARTIF file conforms to the Blind MARTIF DTD. Secondly, the Blind MARTIF file is validated using the validation software described in Appendix C.

1.5.3.2 Is sufficient information preserved?

What are valid results? Since I am not privileged to have all the sample database programs at our disposal, three cross-checks are performed to ensure that the sample import files are indeed valid. First, the produced data files are thoroughly checked by a human to determine if all of the samples are composed of the same data categories as the original data samples. Secondly, four small "format validation" programs are written to verify the results of this check. And finally, samples are sent to an outside, neutral party for evaluation.

At each step, i.e. after each conversion from the sample formats to Blind MARTIF and into each sample format, a human manually checks how much preservation of information there is, exactly what information has been lost, and make a judgement as to whether it is acceptable.

2 Analysis and Algorithms

2.1 Data sample extracts

The most difficult issue to overcome in the analysis and manipulation of the sample data was the manual evaluation of the various data categories. The externally supplied data was provided from Bennett with only minimal descriptions and definitions of the data categories included in each of the data samples. Only through careful research and evaluation of the content associated with each of the data categories was it possible to determine their exact conceptual meaning and significance.

In the following sections of my thesis I describe the data categories associated with each data model, what description was provided by Bennett for each, and what Blind MARTIF data category is used to preserve that information.

2.2 Conversion routines

The conversion routines were written in Visual Basic 5.0 by the coder. Some of the difficulties found in writing the conversion routines were (1) understanding the source data provided by Bennett, (2) determining how best to import the file for conversion, i.e. whether to import the file one section at a time or whether to import the entire file at once for conversion, and (3) converting the resulting Blind MARTIF files back into the sample database formats.

As mentioned in section 3, one suggested improvement for conducting this experiment successfully in the future will be to have access to and carefully analyze each of the sample database format specifications. This will ensure that any possible example, i.e. any valid sample from the sample database, would be easily handled by both Blind MARTIF import and export conversion routines.

2.3 Results

The results of the conversions from SOURCE to Blind MARTIF were almost entirely as expected—all information from the source formats was preserved except for one data category (entry class) that is disallowed (see Appendix E) by Blind MARTIF because it is source termbase specific and three data categories in the Eurodicautom conversion that should have been preserved. Taking into account any data categories that are disallowed by Blind MARTIF and the fact that the conversion programs should improve with future versions, conversions from the sample data formats to Blind MARTIF were able to preserve the information from 5 of 5 fields represented in the GlobeDisk data sample, 19 of 19 of the MultiTerm fields, 17 of 20 of the Eurodicautom fields, and 5 of 5 of NACE fields for a total percentage of preservation of 93.88%. This percentage can and will be increased in the future as new export conversion routines are coded to handle the *custom* data category. This data category will allow incorporation of system-specific data categories in the Blind MARTIF document. This is not to say the use of this data category will enable new import routines to handle the content of such data categories, but it will at a minimum preserve the information to allow a human to recognize and manipulate this information. Further, as the conversion routines become more developed, the more they will be able to properly handle such information.

Difficulties were encountered in determining the exact meaning of the various data categories used in the sample data formats. Without access to the complete specifications of each sample format and having to rely strictly on the explanations of each format given by Bennett, there were instances when the meaning of a particular data category was difficult to determine. There were inconsistencies even within the sample files themselves that made not

only the determination of data category meanings difficult, but also creating conversion routines to handle the inconsistencies difficult. However, through expert knowledge on the part of coder and his programming skills, conversion routines were written to successfully handle the information.

The results of the import conversions from Blind MARTIF, i.e. taking the Blind MARTIF files created from the SOURCE to Blind MARTIF conversions back into the sample formats, were successful. All information was preserved in the resulting 16 documents (see Appendix F) except for the following data categories:

1. Eurodicautom to Blind MARTIF (BTF) to Eurodicautom conversion –The "%%AB" (abbreviation or acronym) is lost. The reason for this loss is due to programming error, rather than a deficiency or inability of Blind MARTIF to handle this particular data category. This, however, was the only instance of information loss among all the "SOURCE to BTF to SOURCE" conversions (e.g., Eurodicautom to BTF to Eurodicautom).

2. Eurodicautom to Blind MARTIF (BTF) to Eurodicautom conversion – "{ECL}" (MultiTerm entry class) is lost. However, this is lost because this is disallowed (see Appendix E) in Blind MARTIF.

The remaining data categories were either all (1) preserved, (2) disallowed in Blind MARTIF (see Appendix E), or (3) lost because the target sample format did not provide equivalent data categories to pass on the information. Again, this information may be preserved by the use of the *custom* data category as described at the beginning of this section.

3 Evaluation and suggestions for further research

3.1 Evaluation

3.1.1 Information loss with Blind MARTIF

Since this experiment clearly illustrates Blind MARTIF capability to preserve the information within the sample database formats, and since the only information lost is either due to programming error or because of ambiguity inherent in the source formats, Blind MARTIF should be considered a valid concept and should be considered by all in the translation industry, more specifically those involved in terminology management, as a vital part of maintaining and exchanging terminology.

3.1.2 Degree of preservation in Blind MARTIF

The degree of preservation afforded by using Blind MARTIF as an interchange format for terminology, between 85% and 100% as described in section 2.5, should be considered acceptable only after others are able to duplicate these results. Every effort has been taken to ensure that as much information is preserved in any such conversions as possible. At a minimum, the information loss is minimal and will most likely be minimized even further in future as continued research is done on how to best preserve the various data categories and how to best program the conversion routines to handle the information.

3.1.3 Possible improvements in Blind MARTIF

The conversion of "Eurodicautom to BTF to Eurodicautom" must necessarily preserve the acronym data category. Such information loss is not acceptable, but can be easily preserved with future conversion routine versions. Unfortunately, there are data base formats that use proximity or linear representation of data for terminology management. For

example, a simple list of terms in a text document may use proximity of a data category to a term to show association of the data category to that term. This can result in ambiguity of meaning, which is often only decipherable by a human.

Conversion routines may take on various forms by allowing for more or less human interaction as desired during conversion. Termbases can be extremely large and automatic conversion (i.e. blind interchange) is naturally advantageous. However, conversions may be improved by allowing the terminology manager to stipulate the handling (possibly software settings) of possible ambiguous or disallowed data categories (e.g. by using a *custom* datCat to preserve information) from both import and export conversion routines. This would still maintain the concept of "blindness", since it is not the purpose of Blind MARTIF to define the meaning of certain, possibly ambiguous sample format data categories, rather to preserve as much of that information as possible and allow import routines (e.g. Blind MARTIF to SOURCE) to handle such information. However, it should be noted that the preservation of such information using a *custom* data category does not guarantee that such information will be preserved in any further conversion.

3.2 Suggestions for further research

In the short term, tests need to be conducted on a much larger scale of real-life data, including the analysis of further termbases not included in this thesis. Further, there needs to be implementation and testing of user subsets as described in section 1.4.3.3, such as TBX (termbase exchange, Lisa/OSCAR subset) and MARTIF Lite. There also need to be an improvement upon the "Blind MARTIF to HTML" viewer. Blind MARTIF would be ideal for storing pools of terminology files on the Internet for people to download or view using a "Blind MARTIF to HTML" viewer. This could be accomplished by creating a browser that

would properly handle a Blind MARTIF document and allow the human to read the file

much as one does using a browser to view HTML on the Internet. Finally, software needs to

be developed to derive a conceptual system from the links in a termbase and software to

check a proposed subset DCS to see if it is truly a subset of the maxi-set.

Appendix A: Formal Specifications of Blind MARTIF
A.1 Graphical representation of a Blind MARTIF document
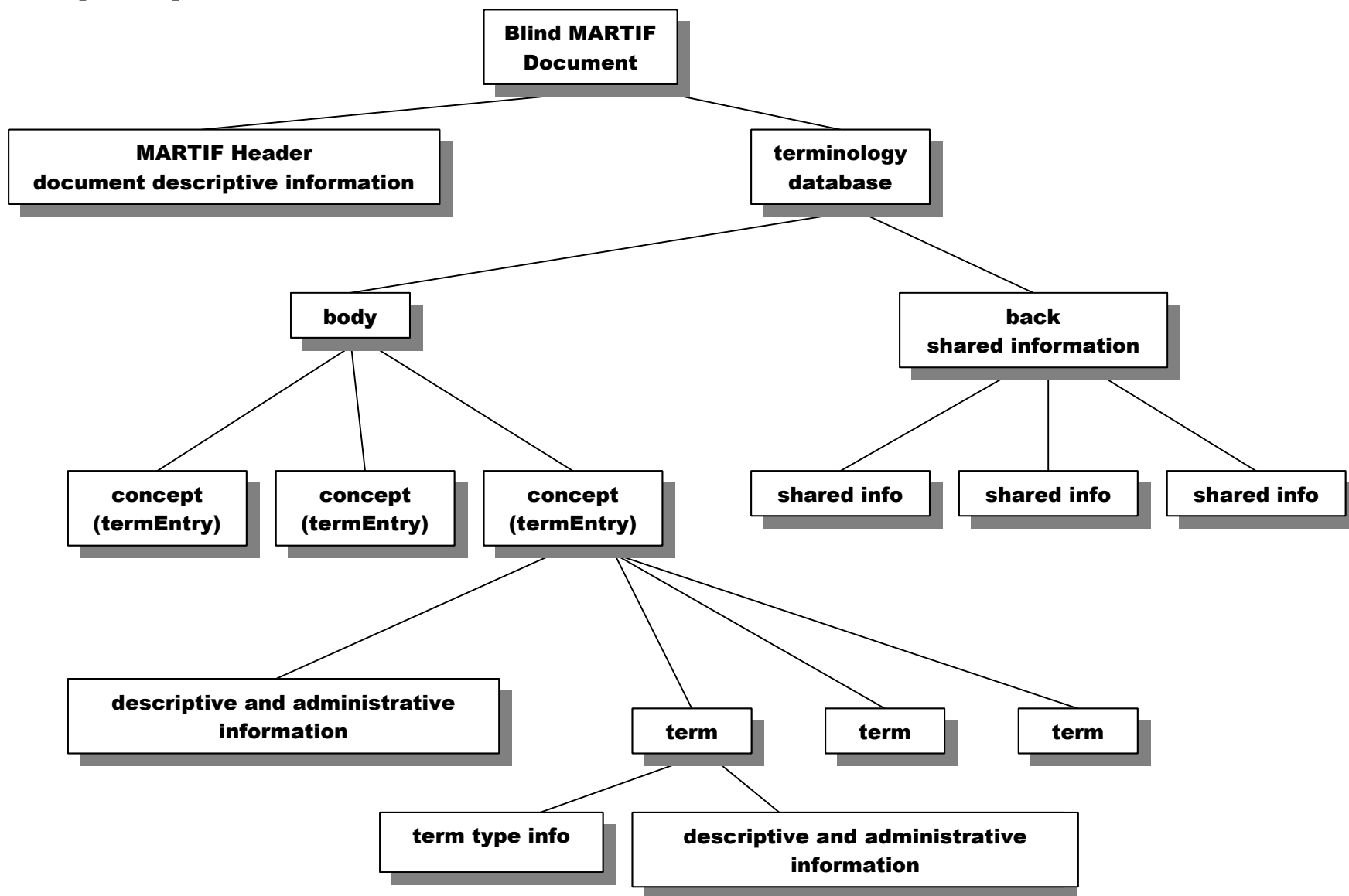
A.2 Sample Blind MARTIF Document

```
<?XML version="1.0"?>
<!DOCTYPE martif PUBLIC "ISO 12200:1998//DTD for MARTIF (Part 2 V1)//EN" [
<!ENTITY % latin1.ents PUBLIC
 'ISO 8879-1986//ENTITIES Added Latin 1//EN' >
%latin1.ents;
<!ENTITY % datcats PUBLIC
 'ISO 12200:1998//ENTITIES for MARTIF data categories (Part 2 V1)//EN'>
]>

<martif type=Part2v1 lang=en>
```

<!-- ***MARTIF Header****: This section contains document specific information, such as (1) a description of the document (i.e., title, publication date, etc.), (2) how the document is encoded, and (3) a revision history of the document. -->*

```
<martifHeader>
<fileDesc>

<titleStmt><title>Example E.1 a complete MARTIF document</title>
</titleStmt>
<publicationStmt><p>unpublished: for TC 37 use only</p></publicationStmt>
<sourceDesc><p>from ISO DIS 12200</p></sourceDesc>

</fileDesc>

<encodingDesc>

<p type=datCatSetVersion>1</p>
<p type=languageDeclaration id=en>English</p>
<p type=languageDeclaration id=fr>French</p>
<p type=languageDeclaration id=de>German</p>
</encodingDesc>

<revisionDesc>
<change><p>1995-02-04: extracted from 12200 wp file</p></change>
<change><p>1997-02-26: adjusted for blind DTD</p></change>
<change><p>1997-10-28: added slash to ptr tag (for XML)</p></change>
<change><p>1997-11-04: added type attrib to martif element</p></change>
<change><p>1998-01-20: tweaked file into compliance w/ newest blind
DTD</p></change>
</revisionDesc>
</martifHeader>

<text>
```

<!-- ***Body****: This section contains the actual concept entries and the terms that are associated with each concept. As Melby (1998) points out, it should be noted that each concept entry may have many language sections that contain one or more terms. -->*

```
<body>
```

```
<termEntry>
```

<!-- ***Descriptive and administrative information****: This information at the beginning of the concept entry pertains to the entire entry.* -->

```
<ref type='subjectField' target='mtlAppearance'> appearance of materials
</ref>
<note> The in-house working group for Optics is slated to finalize this
entry by 1995-12-15. </note>
```

<!-- ***langSet****: As described above, this concept entry has a number of language sections, in this case the first being English.* -->

```
<langSet lang=en>

<ntig>
```

<!-- ***Term****: The term group may contain notes such as part of speech and gender (term type info).* -->

```
<termGrp>
<term> opacity </term>
<termNote type='partOfSpeech'> noun </termNote>
</termGrp>

<descripGrp>
<descrip type='definition'> degree of obstruction to the transmission of
visible
light </descrip>
<ptr type='sourceIdentifier' target='ASTM.E284'/>
</descripGrp>
```

<!-- ***ptr****: Each term may point to shared information in the back of the MARTIF document..*
 -->

```
<ptr type='figure' target='f357'/>
<ptr type='responsibility' target='ASTM.E12'/>
<date type='origination'>1995-02-04</date>
</ntig>

</langSet>

<langSet lang=de>

<ntig>

<termGrp>
<term> Opazit&#x00E4;t </term>
<termNote type='partOfSpeech'> noun </termNote>
<termNote type='grammaticalGender'> feminine </termNote>
</termGrp>

<descripGrp>
<descrip type='definition'> Ma&#x00DF; f&#x00FC;r die
Lichtundurchl&#x00E4;ssigkeit </descrip>
```

```
<ref type='sourceIdentifier' target='DIN-6730-1992-08'> p. 5 </ref>
</descripGrp>
<ptr type='responsibility' target='NPa'/>
</ntig>

</langSet>

<langSet lang=fr>

<ntig>

<termGrp>
<term> opacit&#x00E9; </term>
<termNote type='partOfSpeech'> noun </termNote>
<termNote type='grammaticalGender'> feminine </termNote>
</termGrp>

<descripGrp>
<descrip type='definition'> rapport du flux lumineux incident au flux
lumineux transmis ou
r&#x00E9;fl&#x00E9;chi par un noircissement photographique </descrip>
<ptr type='sourceIdentifier' target='JHdi1986'/>
</descripGrp>

<ptr type='responsibility' target='CIRAD'/>
</ntig>

</langSet>

</termEntry>

</body>
```

<!-- ***back****: As shown in the graphical representation of the high-level structure, the back
contains information that is shared among concepts. Among other advantages of pointing to
shared information is that in doing so, the database manager can encourage consistency
within the document, i.e., reducing the possibility of error by eliminating the need to enter
the same shared information for each associated term*-->

```
<back>

<refObjectList type=subjectFieldSet>
 <refObject id='mtlAppearance'>
  <item type='subjectFieldName'> appearance of materials </item>
  <item type='description'> treated in DIN under paper and cardboard
</item>
 </refObject>
</refObjectList>

<refObjectList type=responsibleParty>
 <refObject id='ASTM.E12'>
  <item type='organizationName'>ASTM</item>
  <item type='fname'>John</item>
  <item type='surname'>Doe</item>
 </refObject>
```

48

```
 <refObject id='NPa'>
  <item type='organizationName'>Normenaussch&#x00DF; Papier und
Pappe (NPa) im DIN Deutsches Institut f&#x00FC;r Normung e.V.</item>
 </refObject>

 <refObject id='CIRAD'>
  <item type='organizationName'>CIRAD (Conseil international de la
langue fran&#x00E7;aise  Vocabulaire de la radiographie</item>
 </refObject>
</refObjectList>

<refObjectList type=binaryData>
 <refObject ID='f357'>
  <item type='name'>Degrees of Opacity</item>
  <!-- bundled graphic to be inserted here -->
  <item type='description'> graphics chart illustrating degrees of
opacity</item>
 </refObject>
</refObjectList>

<refObjectList type=bibl>
 <refObject id='ASTM.E284'>
  <item type='title'> Terminology of Appearance </item>
  <itemSet type='author'>
   <item type='corpauth'> ASTM TC E 12 </item>
   <item type='role'> technical committee </item>
   <item type='publisher'> ASTM </item>
  </itemSet>
  <itemSet>
   <item type='city'> Philadelphia </item>
   <item type='country'> U.S.A.</item>
  </itemSet>
  <item type='publicationDate'> 1990 </item>
 </refObject>

 <refObject id='DIN-6730-1992-08'>
  <item type='category'> national standard </item>
  <item type='title'> Papier und Pappe -- Begriffe </item>
  <item type='author'>  Normenaussch&#x00DF; Papier und Pappe (NPa)
im DIN Deutsches Institut f&#x00FC;r Normung e.V. </item>
  <item type='publisher'> Beuth Verlag </item>
  <itemSet>
   <item type='city'> Berlin </item>
   <item type='country'> Germany </item>
  </itemSet>
  <item type='publicationDate'> 1983 </item>
 </refObject>

 <refObject id='JHdi1986'>
  <item type='title'> Dictionnaire des industries: 36000 definitions index
anglais-fran&#x00E7;ais </item>
  <item type='author'> CIRAD (Conseil international de la langue
fran&#x00E7;aise
Vocabulaire de la radiographie </item>
  <item type='publisher'> Conseil international de la langue
fran&#x00E7;aise </item>
  <item type='city'> Paris </item>
```

```
  <item type='publicationDate'> 1986 </item>
 </refObject>

</refObjectList>

</back>

</text>

</martif>
```

A.3 Blind MARTIF Data Categories, Maximal Set

The data categories found in the Blind MARTIF data category maximal set (maxi-set) are taken from ISO FDIS 12620. This should provide the maxi-set with a wide base of data categories, representative of data categories found in most terminology databases.

datCat Name	GI	12620 Section
Class — Term		
Section 1		
term	term	A.1
Class – Term-related Information		
Section 2 — Properties of a Term		
term type	termNote	A.2.1…
part of speech	termNote	A.2.2.1
grammatical gender	termNote	A.2.2.2
grammatical number	termNote	A.2.2.3
animacy	termNote	A.2.2.4
usage note	termNote	A.2.3.1
geographical usage	termNote	A.2.3.2
register	termNote	A.2.3.3
frequency	termNote	A.2.3.4
temporal qualifier	termNote	A.2.3.5
time restriction	termNote	A.2.3.6
proprietary restriction	termNote	A.2.3.7
term provenance	termNote	A.2.4.1
etymology	termNote	A.2.4.2
pronunciation	termNote	A.2.5
syllabification	termNote	A.2.6
hyphenation	termNote	A.2.7
morphological element	termNote	A.2.8.1
term element	termNote	A.2.8.2
normative authorization	termNote	A.2.9.1
language-planning qualifier	termNote	A.2.9.2
Section 3 — Equivalency		
false friend	termNote	A.3.2
false friend	ref	A.3.2
reliability code	termNote	A.3.4
transfer comment	termNote	A.3.5

| transfer comment | ref | A.2.5 |
| | | |

Class — Descriptive Information		
Section 4 — Subject Field & Classification Systems		
subject field	ref	A.4
classification number	ref	A.4.1-A.4.2

Section 5 — Definitions, etc.		
definition	descrip	A.5.1
explanation	descrip	A.5.2
context	descrip	A.5.3
example	descrip	A.5.4
figure	ptr	A.5.5.1
audio	ptr	A.5.5.2
video	ptr	A.5.5.3
table	ptr	A.5.5.4
other Binary Data	ptr	A.5.5.5
unit	descrip	A.5.6
range	descrip	A.5.7
characteristic	descrip	A.5.8

| **Section 6: subsumed into Section A.7.1/2** | | |
| | | |

Section 7 — Concept Systems		
concept position	ref	A.7.2
broader concept generic	ref	A.7.2.1
broader concept partitive	ref	A.7.2.1
superordinate concept generic	ref	A.7.2.2
superordinate concept partitive	ref	A.7.2.2
subordinate concept generic	ref	A.7.2.3
subordinate concept partitive	ref	A.7.2.3
coordinate concept generic	ref	A.7.2.4
coordinate concept partitive	ref	A.7.2.4
related concept	ref	A.7.2.5

Section 8 — Miscellaneous		
note		A.8
description note		A.11.3.1*
administrative note		A.11.3.3*
* Variation on <note> A.8.		

Class — Administrative Information		
Section 9 — Documentary Languages & Thesauri		
thesaurus descriptor	ref	A.9.1-A.9.2
top term	ref	A.9.2.1
broader term	ref	A.9.2.2
narrower term	ref	A.9.2.3
related term	ref	A.9.2.4
keyword	admin	A.9.4
index heading	admin	A.9.5
Section 10 — Other Administrative Information		
origination	date	A.10.2.1.1
input	date	A.10.2.1.2
modification	date	A.10.2.1.3
check	date	A.10.2.1.4
approval	date	A.10.2.1.5
withdrawal	date	A.10.2.1.6
standardization	date	A.10.2.1.7
exportation	date	A.10.2.1.8
importation	date	A.10.2.1.9
responsibility	ptr	A.10.2.2
originator	ptr	A.10.2.2.1
inputter	ptr	A.10.2.2.2
updater	ptr	A.10.2.2.3
checker	ptr	A.10.2.2.4
approver	ptr	A.10.2.2.5
user	ptr	A.10.2.2.6
withdrawer	ptr	A.10.2.2.7
exporter	ptr	A.10.2.2.8
importer	ptr	A.10.2.2.9
subset owner	ptr	A.10.2.2.10
customer subset	ref	A.10.3.1
initial customer subset	ref	A.10.3.2
project subset	ref	A.10.3.3
initial project subset	ref	A.10.3.4
product subset	ref	A.10.3.5
application subset	ref	A.10.3.6
environment subset	ref	A.10.3.7
business unit subset	ref	A.10.3.8
security subset	admin	A.10.3.9
user suggestion	admin	A.10.5

entailed term	hi	A.10.6.1
sort key	admin	A.10.6.2
search term	admin	A.10.6.3
foreign	foreign	A.10.8
cross-reference	ptr	A.10.18
cross-reference	ref	A.10.18
homograph	ref	A.10.18.5
antonym	ref	A.10.18.6
homograph	admin	A.10.18.5
antonym	admin	A.10.18.6
source identifier	ptr	A.10.20
source identifier	ref	A.10.20
URL	ptr	A.10.21.1
FPI	ptr	A.10.21.2
originating institution	ptr	A.10.22.2
maxi	ref	
Custom	ref	

A.4 Core DTD of Blind MARTIF

```
<!--

======================================================================
Preliminary Blind MARTIF Version 1 DTD (MTFP2V1.DTD)
======================================================================


TYPICAL DOCTYPE DECLARATION:

<!DOCTYPE martif PUBLIC "ISO 12200:1998//DTD for MARTIF (Part 2 V1)//EN" [
<!ENTITY % datcats PUBLIC
 'ISO 12200:1998//ENTITIES for MARTIF data categories (Part 2 V1)//EN'>
]>

Note: Before the declaration of the datcats entity, there
may be the incusion of one or more character entity sets.
For example, the Latin1 entities would be included as follows:

<!ENTITY % latin1.ents PUBLIC
 'ISO 8879-1986//ENTITIES Added Latin 1//EN' >
%latin1.ents;


<!-- ORIGIN AND HIGH-LEVEL STRUCTURE


======================================================================


Submitted 1997-10-11 (to TC 37 project group, including Bennett and LeMeur)

Partially aligned with February 1997 version of MARTIF Part 1 DTD.

BASED ON DTD in Hardman thesis of fall 1996
 (which was based on the 1995 version of MARTIF)

(c) 1998 BYU TRG and Daniel Hardman
(until officially submitted as part of an ISO TC 37 work item proposal)

High-level structure of a Blind MARTIF document

martif == (martifHeader, text)
 martifHeader == (fileDesc, encodingDesc, revisionDesc?)
 text == (body, back?)
  body == (termEntry+)
  back == ((refObjectList)*)
   refObjectList == ((refObject)+)
    refObject ==  ((itemSet | itemGrp | item)+)

Structure of a concept-oriented entry (a termEntry)

 termEntry == ((%auxInfo;),(langSet+))
  langSet == ((%auxInfo;), (ntig)+)
   ntig == (termGrp, %auxInfo;)
    termGrp == (term, (termNote | termNoteGrp)*, %noteLinkInfo)
     termNoteGrp == (termNote, %noteLinkInfo;)

For more detail on unexpanded elements, consult
alphabetical listing of all elements (below):
```

```
-->


<!--
===============================================================================
SOME USEFUL ENTITIES THAT ARE REFERENCED BELOW
===============================================================================
-->

<!ENTITY % basicText '(#PCDATA | hi)*' >
<!ENTITY % noteText '(#PCDATA | hi | foreign)*' >

<!ENTITY % auxInfo '(descrip | descripGrp | admin | adminGrp |
             ptr | ref | date | note)*' >
<!ENTITY % noteLinkInfo '(ptr | ref | date | note)*' >
<!ENTITY % dNoteLinkInfo '(ptr | ref | date | note | descripNote)*' >
<!ENTITY % aNoteLinkInfo '(ptr | ref | date | note | adminNote)*' >

<!-- Entities that define attributes -->
<!ENTITY % impID 'id ID #IMPLIED'>
<!ENTITY % impIDLang 'id ID #IMPLIED lang IDREF #IMPLIED' >
<!ENTITY % impIDType 'id ID #IMPLIED type CDATA #IMPLIED' >
<!ENTITY % impIDLangType '%impIDType lang IDREF #IMPLIED' >
<!ENTITY % reqID 'id ID #REQUIRED' >
<!ENTITY % reqLang.impID 'id ID #IMPLIED lang IDREF #REQUIRED' >
<!ENTITY % reqType.impID 'id ID #IMPLIED type CDATA #REQUIRED' >
<!ENTITY % reqLangType.impID '%reqLang.impID type CDATA #REQUIRED' >


<!--
===============================================================================
ELEMENTS (IN ALPHABETICAL ORDER)
===============================================================================
-->
<!ELEMENT admin  (%basicText;) >
<!ELEMENT adminGrp  (admin, %aNoteLinkInfo;) >
<!ELEMENT adminNote  (%noteText;) >
<!ELEMENT back  ((refObjectList)*) >
<!ELEMENT body  (termEntry+) >
<!ELEMENT change  (p+) >
<!ELEMENT date (#PCDATA) >
<!ELEMENT descrip  (%noteText;) >
<!ELEMENT descripGrp  (descrip, %dNoteLinkInfo; ) >
<!ELEMENT descripNote  (%noteText;) >
<!ELEMENT encodingDesc  (p+) >
<!ELEMENT fileDesc  (titleStmt, publicationStmt, sourceDesc+) >
<!ELEMENT foreign  (%basicText;) >
<!ELEMENT hi  (#PCDATA) >
<!ELEMENT item  (%basicText) >
<!ELEMENT itemGrp  (item, %noteLinkInfo;)>
<!ELEMENT itemSet  ((item | itemGrp)+)>
<!ELEMENT langSet  ((%auxInfo;), (ntig)+) >
<!ELEMENT martif  (martifHeader, text) >
<!ELEMENT martifHeader  (fileDesc, encodingDesc, revisionDesc?) >
```

```
<!ELEMENT name  (#PCDATA) >
<!ELEMENT note  (%noteText;) >
<!ELEMENT ntig  (termGrp, %auxInfo;) >
<!ELEMENT p  (%noteText;) >
<!ELEMENT ptr  EMPTY >
<!ELEMENT publicationStmt  (p+) >
<!ELEMENT ref  (#PCDATA) >
<!ELEMENT refObject  ((itemSet | itemGrp | item)+ | (xref, note?)) >
<!ELEMENT refObjectList  ((refobject)+) >
<!ELEMENT resp  (name, role) >
<!ELEMENT revisionDesc  (change+) >
<!ELEMENT role  (#PCDATA) >
<!ELEMENT sourceDesc  (p+) >
<!ELEMENT term  (%basicText;) >
<!ELEMENT termEntry  ((%auxInfo;),(langSet+)) >
<!ELEMENT termGrp  (term, (termNote | termNoteGrp)*, %noteLinkInfo) >
<!ELEMENT termNote  (%noteText;) >
<!ELEMENT termNoteGrp  (termNote, %noteLinkInfo;) >
<!ELEMENT text  (body, back?) >
<!ELEMENT title  (#PCDATA) >
<!ELEMENT titleStmt  (title, resp*) >
<!ELEMENT xref  (#PCDATA) >

<!--
======================================================================
ATTRIBUTES THAT APPLY TO EACH ELEMENT
======================================================================
-->


<!-- include external attribute declarations -- DCE File -->
%datcats;

<!--     admin: type attribute declared externally in datcats -->
<!ATTLIST adminGrp id IDREF #IMPLIED >
<!ATTLIST adminNote %impIDLang; >
<!ATTLIST back id IDREF #IMPLIED >
<!ATTLIST change %impIDLang; >
<!--     date: type attribute declared externally in datcats -->
<!--     descrip: type attribute declared externally in datcats -->
<!ATTLIST descripGrp id IDREF #IMPLIED >
<!ATTLIST descripNote %impIDLang; >
<!ATTLIST encodingDesc id IDREF #IMPLIED >
<!ATTLIST fileDesc id IDREF #IMPLIED >
<!ATTLIST foreign %reqLang.impID; >
<!--     hi: type attribute declared externally in datcats -->
<!ATTLIST item %impIDType; >
<!ATTLIST itemGrp %impIDType; >
<!ATTLIST itemSet %impIDType; >
<!ATTLIST langSet %reqLang.impID; >
<!ATTLIST martif %reqLangType.impID; >
<!ATTLIST martifHeader %impIDType; >
<!ATTLIST note %impIDLang; >
<!ATTLIST ntig %impID; >
<!ATTLIST p %impIDLangType; >
<!--     ptr: type attribute declared externally in datcats -->
```

```
<!ATTLIST publicationStmt %impID; >
<!--    ref: type attribute declared externally in datcats -->
<!ATTLIST refObject %impID; >
<!ATTLIST refObjectList %reqType.impID; >
<!ATTLIST resp %impID; >
<!ATTLIST revisionDesc %impIDLang; >
<!ATTLIST role %impIDLang; >
<!ATTLIST sourceDesc %impIDLang; >
<!ATTLIST term %impID; >
<!ATTLIST termEntry %impID; >
<!ATTLIST termGrp %impID; >
<!--    termNote: type attribute declared externally in datcats -->
<!ATTLIST termNoteGrp %impID; >
<!ATTLIST text %impID; >
<!ATTLIST title %impIDLang; >
<!ATTLIST titleStmt %impIDLang; >
<!ATTLIST xref %impID; target CDATA #REQUIRED >
<!-- end -->
```

A.5 DCE—a logical Part of the Blind MARTIF DTD

```
<!ATTLIST admin

 TYPE
 (
  keyword|
  indexHeading|
  securitySubset|
  userSuggestion|
  sortKey|
  searchTerm|
  homograph|
  antonym
 )
 #REQUIRED
>

<!ATTLIST date

 TYPE
 (
  origination|
  input|
  modification|
  check|
  approval|
  withdrawal|
  standardization|
  exportation|
  importation
 )
 #REQUIRED
>

<!ATTLIST descrip

 TYPE
 (
  definition|
  explanation|
  context|
  example|
  unit|
  range|
  characteristic
 )
 #REQUIRED
>

<!ATTLIST ptr

 TYPE
 (
  figure|
```

```
 audio|
 video|
 table|
 otherBinaryData|
 responsibility|
 originator|
 inputter|
 updater|
 checker|
 approver|
 user|
 withdrawer|
 exporter|
 importer|
 subsetOwner|
 crossReference|
 sourceIdentifier|
 URL|
 FPI|
 originatingInstitution
 )
 #REQUIRED

 TARGET IDREF #REQUIRED
>

<!ATTLIST ref

 TYPE
 (
 falseFriend|
 transferComment|
 subjectField|
 classificationNumber|
 conceptPosition|
 broaderConceptGeneric|
 broaderConceptPartitive|
 superordinateConceptGeneric|
 superordinateConceptPartitive|
 subordinateConceptGeneric|
 subordinateConceptPartitive|
 coordinateConceptGeneric|
 coordinateConceptPartitive|
 relatedConcept|
 thesaurusDescriptor|
 topTerm|
 broaderTerm|
 narrowerTerm|
 relatedTerm|
 customerSubset|
 initialCustomerSubset|
 projectSubset|
 initialProjectSubset|
 productSubset|
 applicationSubset|
 environmentSubset|
```

```
  businessUnitSubset|
  crossReference|
  homograph|
  antonym|
  sourceIdentifier|
  custom|
  maxi
 )
 #REQUIRED
 TARGET IDREF #REQUIRED
>

<!ATTLIST termNote

 TYPE
 (
 termType|
 partOfSpeech|
 grammaticalGender|
 grammaticalNumber|
 animacy|
 usageNote|
 geographicalUsage|
 register|
 frequency|
 temporalQualifier|
 timeRestriction|
 proprietaryRestriction|
 termProvenance|
 etymology|
 pronunciation|
 syllabification|
 hyphenation|
 morphologicalElement|
 termElement|
 normativeAuthorization|
 languagePlanningQualifier|
 falseFriend|
 reliabilityCode|
 transferComment
 )
 #REQUIRED
>
```

Appendix B: Externally supplied test data

B.1 Sample GlobeDisk (GBD) data

(NOTE: The GlobeDisk data received from Bennett contained a few flaws, where certain characters were replaced. For example, the "ó" character was replaced by "½". This appears to be a flaw with some conversion routine applied to this data before the data was delivered from Bennett.)

##13ENES##,LTIMA VALIDACI…N #23.02.96European Cooperative Research Action for Technology (CRAFT) [IDT]#[13/94. Este programa tiene por objetivo aportar apoyo financiero a grupos de empresas industriales -principalmente PYME-, unidas por la necesidad comœn de seguir haciendo investigaci½n industrial o tecnol½gica, que les permita asociarse y encargar a terceros (centros de investigaci½n, universidades u otras empresas) la ejecuci½n de un contrato de investigaci½n y desarrollo tecnol½gico (IDT) para ellas. (doc. XXIII/94/ 05200000.p00)] Programa europeo de investigaci½n tecnol½gica en cooperativa (CRAFT) [JC]

push-pull activity [VALUE]#[13/14490000/94] actividad de empuje-tir½n [JC]
R&D Limited Partnerships [INNOVATION]   # [13/95/53650000.P00] Sociedades Limitadas de I+D [ar]
task force [VALUE]#[13/14520000.p00] grupo operativo [JC]
awareness [IMPACT]#[13/] concienciaci½n [AQ];[1257/94. Promotion of awareness in the EEA. (Anexo IV)] conocimiento [JC];[Sensibilizaci½n en cuestiones medioambientales.] sensibilizaci½n [AQ]
back-up initiative [VALUE]#[13/14490000/94] iniciativa de reserva [JC]
benchmarking#[13/95/50660000.P00] patrones de referencia (establecimiento de) [candemi]
bottom-up approach [VALUE]#[13/14490000/94] enfoque ascendente [JC]
business plan [VALUE]#[13/14490000/94] plan de negocio [JC]
Central Support Team (CST) [IMPACT]#[13/1257/94. Anexo I] Equipo central de apoyo (CST) [JC]
Community Innovation Survey (CIS) # [13/] Encuesta de la Innovaci½n Comunitaria [aq]
context driven application [VALUE]#[13/14490000/94] aplicaci½n por contexto [JC]
demand pull [VALUE]#[13/14490000/94. Una innovaci½n o invento se desarrolla en funci½n de las exigencias de la demanda (concepto contrario al de Technology Push)] tir½n de la demanda [JC]
dissemination [VALUE]#[13/14490000.p00. Explotaci½n y difusi½n de los resultados de investigaci½n.] difusi½n [JC]
distribution network [VALUE]#[13/14490000/94] red de distribuci½n [JC]
European Association for the Transfer of Technology, Innovation and Industrial Information (TII) [SPRINT]#[13/Puede aparecer la traducci½n de la sigla TII como TECNOLOG™A, INNOVACI…N, INFORMACI…N] Asociaci½n Europea de Transferencia de Tecnolog-a, Innovaci½n e Informaci½n Industrial (TII) [JC]
European Commission Host Organization (ECHO) [IMPACT]#[13] Organizaci½n de distribuci½n de la Comisi½n Europea (ECHO) [JC]
European Innovation Monitoring System (EIMS)#[13/94/16440400.P00] Observatorio Europeo de la Innovaci½n (EIMS), traducci½n propuesta por Jos_ Ram½n Tiscar [mt]

European Securities Dealers Association [INNOVATION]   # [13/95/53650000.P00]
Asociaci½n Europea de Intermediarios Financieros [ar]
event#[13/1257\94. Anexo I (seminars, workshops, exhibitions, conferences..)] acto [JC]
event schedule [contratros]#[13/1257/94] programa de actividades [JC]
ex ante [VALUE]#[13/14490000/94: integraci½n previa o ex ante de las perspectivas
empresariales en los programas de IDT] ex ante [JC]
exploitation [VALUE]#[13/14490000/94. Explotaci½n y difusi½n de los resultados de
investigaci½n.] explotaci½n [JC]
hardware [inform tica]#[13] equipo(s) inform tico(s)
Help Desk [VALUE]#[13/14490000/94] Servicio de asistencia t_cnica [JC]
host [VALUE]#[13/14490000/94.Organismo, entidad, etc., como una librer-a, que acoge un
servicio (por ejemplo, los centros de enlace VALUE)] organismo anfitri½n [JC]
host [inform tica]#[13/] distribuidor [JC]
Innovation Relay Centres (IRC) [difusi½n de resultados]   # [13/94/17870100.P00. Centros
asesores de orientaci½n pr ctica que facilitan a sus clientes el acceso a la investigaci½n y la
tecnolog-a. Su objetivo principal es promocionar la transferencia de resultados de la
investigaci½n y de tecnolog-as de conformidad con las necesidades manifestadas por el
tejido industrial local, con objeto de mejorar su competitividad a trav_s de la innovaci½n.]
centros de enlace de la innovaci½n (IRC) [JC]

B.1.1 GlobeDisk to Blind MARTIF Mapping

Field	Description	Examples	12620 datCat
##…#	EN-ES terminology for the Commission's DG13	##13ENES#	\<langSet lang='…'\> \<p type=languageDeclaration ID='…'\>
#ULTIMA VALIDACI…N #DD.MM.YY	Last date of validation	#ULTIMA VALIDACI…N #23.02.96	\<revisionDesc\> Perhaps attach this date to each concept entry.
#	Separates the source from the target	N/A	N/A – Conversion algorithm would identify all text outside of square brackets on the left side of the # as the term in English (source). All text outside of square brackets on the right side is the term in Spanish (target).
[…] (left side)	Typically indicates context (optional) Through evaluation with Ryan and Alan, this appears to be the SUBJECT field.	[VALUE] [INNOVATION] [IMPACT] [SPRINT] [contratros] [inform tica]	\<note\> This is a difficult one as well. It is my opinion that there is no structure to this category—any text is permissible inside the square brackets.
[…] right side, immediately following #	Source (many inconsistencies)	[13/95/53650000.P00] [13/94…(text)…] [13/14490000/94] [13/] 13/14520000.p00] [13/…(text)…] [13] [Doc. es/13/94/16120100.p00]	\<note\> Left as \<note\> as a result of the inconsistencies within this field between entries in this data sample.
[…] right side, last square brackets of target-language term	Typically originator	[JC] [ar] [candemi]	\<ptr type='responsibility' target='…'\> \<ptr type='originator' target='…'\>
;(space)	Separates the target-language terms		

B.2 Sample Eurodicautom (ECO) data

```
%%BE BTL
%%TY MTW96
%%NI 3
%%CF 2
%%AU hames
%%CM ME
%%DE
%%VE Gesamt-Aerosole [1]
%%RF Dok. 0394/95de [VE1]
%%FR
%%VE concentration atmosphérique totale [1]
%%NT {NTE} aérosol solide [VE1]; {CRD} 24.05.1996 - 16:12:02; {CRU}
agl03@hamesae; {CHD} 18.06.1996 - 15:05:12; {CHU} agl03@hamesae; {ECL} 5; {ENR}
3; {DOM} medicine ME; {DOC} Dok. 0394/95de; {TGT} FR
%%RF R. Lauwerys, Toxicologie industrielle, p.10 [VE1]
%%BE BTL
%%TY MTW96
%%NI 4
%%CF 2
%%AU hames
%%CM ME
%%DE
%%VE Aerosol-Durchmesser [1]
%%RF Dok. 0394/95 [VE1]
%%FR
%%VE diamètre des particules de l'aérosol [1]
%%NT {NTE} diamètre aérodynamique des particules [VE1]; {CRD} 24.05.1996 -
16:17:03; {CRU} agl03@hamesae; {CHD} 18.06.1996 - 15:06:45; {CHU} agl03@hamesae;
{ECL} 5; {ENR} 4; {DOM} medicine ME; {DOC} Dok. 0394/95de; {TGT} FR
%%RF R. Lauwerys, Toxicologie industrielle, p.16 [VE1]
%%BE BTL
%%TY MTW96
%%NI 6
%%CF 2
%%AU hames
%%CM AG ST
%%EN
%%VE objective yield model [1]
%%RF OS/1665/95en [VE1]
%%FR
%%VE modèle de prévision de rendements objectifs [1]
%%NT {NTE} prévision des rendements de maïs et de soja aux USA [VE1]; {CRD}
28.05.1996 - 16:49:25; {CRU} agl03@hamesae; {CHD} 18.06.1996 - 15:13:06; {CHU}
```

agl03@hamesae; {ECL} 5; {ENR} 6; {DOM} agriculture AG, statistics ST; {DOC}
OS/1665/95en; {TGT} FR
%%RF OS/demandeur [VE1]
%%BE BTL
%%TY MTW96
%%NI 7
%%CF 2
%%AU hames
%%CM AG ST
%%EN
%%VE area frame [1]
%%RF OS/1665/95en [VE1]
%%FR
%%VE base de sondage aréolaire [1]
%%NT {NTE} prévision des rendements de maïs et de soja aux USA [VE1]; {CRD}
28.05.1996 - 16:54:46; {CRU} agl03@hamesae; {CHD} 18.06.1996 - 15:02:19; {CHU}
agl03@hamesae; {ECL} 5; {ENR} 7; {DOM} agriculture AG, statistics ST; {DOC}
OS/1665/95en; {TGT} FR
%%RF OS/demandeur [VE1]
%%BE BTL
%%TY MTW96
%%NI 8
%%CF 2
%%AU hames
%%CM AG ST
%%EN
%%VE list frame samples [1]; list operations [2]
%%RF OS/1665/95en [VE1]
%%FR
%%VE échantillons sur liste [1]
%%NT {NTE} prévision des rendements de maïs et de soja aux USA [VE1]; {CRD}
28.05.1996 - 16:58:49; {CRU} agl03@hamesae; {CHD} 18.06.1996 - 15:21:26; {CHU}
agl03@hamesae; {ECL} 5; {ENR} 8; {DOM} agriculture AG, statistics ST; {DOC}
OS/1665/95en; {TGT} FR
%%RF OS/demandeur [VE1]
%%BE BTL
%%TY MTW96
%%NI 9
%%CF 2
%%AU hames
%%CM AG ST
%%EN
%%VE farm report panel survey [1]
%%RF OS/1665/95en [VE1]
%%FR
%%VE enquête réalisée auprès d'un panel représentatif d'exploitants [1]

%%NT {NTE} prévision des rendements de maïs et de soja aux USA [VE1]; {CRD} 28.05.1996 - 17:02:42; {CRU} agl03@hamesae; {CHD} 02.09.1996 - 15:14:38; {CHU} agl03@reisant; {ECL} 5; {ENR} 9; {DOM} agriculture AG, statistics ST; {DOC} OS/1665/95en; {TGT} FR
%%RF OS/demandeur [VE1]
%%BE BTL
%%TY MTW96
%%NI 10
%%CF 2
%%AU hames
%%CM AG ST
%%EN
%%VE objective yield forecasting model [1]
%%RF OS/1665/95en [VE1]
%%FR
%%VE modèle de prévision de rendement objectif [1]
%%NT {NTE} prévision des rendements de maïs et de soja aux USA [VE1]; {CRD} 28.05.1996 - 17:07:02; {CRU} agl03@hamesae; {CHD} 18.06.1996 - 15:12:50; {CHU} agl03@hamesae; {ECL} 5; {ENR} 10; {DOM} agriculture AG, statistics ST; {DOC} OS/1665/95en; {TGT} FR
%%RF OS/demandeur [VE1]

B.2.1 Eurodicautom to Blind MARTIF Mapping

Field	Description	Examples	12620 datCat
%%BE	Indicates originating body	%%BE BTL	<ptr type='originatingInstitution' target='…'>
%%TY	"collection" code	%%TY MTW96	<ptr type=X target=Y>
%%NI	Entry Number	%%NI 3	<termEntry id='…'>
%%CF	Reliability code	%%CF 2	<termNote type='reliabilityCode'>
%%AU	Originating terminologist • optional	%%AU hames	<ptr type='originator' target=Y/>
%%CM	One or more two or three letter **Lennoch** subject code • optional	%%CM ME	<ptr type='subjectField target='…'>
%%DE*	Language code * This data category will change with the respective language, e.g. %%FR, etc.	%%DE	<langSet lang='…'>
%%VE	Term	%%VE Gesamt-Aerosole [1]	<term>
%%RF	Reference	%%RF Dok. 0394/95de [VE1]	<ref type='sourceIdentifier' target='…'>
%%NT {NTE}	Note	%%NT {NTE} prévision des rendements de maïs et de soja aux USA [VE1];	<termNote type=termType>TransferComment</termNote>
%%AB	Abbreviation or acronym	%%AB RDT [1]	<termNote type='termType'>abbreviation</termNote>

{CRD}	Entry creation date and time (coded using European date conventions).		`<date type='origination'>...</date>`
{CRU}	The unique UserId of the terminologist creating the entry.		`<ptr type='originator' target=Y/>`
{CHD)	The date and time of the most recent change to the entry.		`<date type='modification'>...</date>`
{CHU}	The unique UserId of the terminologist originating the most recent change to the entry.		`<ptr type='updater' target=Y/>`
{ECL}	The MultiTerm entry class of the relevant record (between 1 and 8), 1 being the lowest (default) class.		Disallowed in Blind MARTIF, although it may affect the process of the export routine, i.e. the entry class may appear in the body, shared references, etc.
{ENR}	The MultiTerm entry number of the relevant record. Normally, this should be the same as the NI (see above).		`<termEntry id='…'>`
{DOM}	Subject codes, accompanied by a short text description of the significance of the code(s).		`<ptr type='subjectField' target='…'>` Already used above
{DOC}	Source document as indicated (as an entry-level attribute) in the MultiTerm entry.		`<ref type=sourceIdentifier target=Z>...</ref>` or `<ptr type=sourceIdentifier target=Z/>`
{TGT}	Target language as indicated in MultiTerm.		`<langSet lang='…'>`

B.3 Sample MultiTerm (MMI/MME) data

**
<Created By>g04
<SourceDoc&Lang>EN/13/1257/94
<Personal Comment>GlobeDisk BTBABR930000390
<EN>CST
<TermTyp>Abrev.
<EN>Central Support Team
<TermTyp>Term
<Note>IMPACT
<ES>equipo central de apoyo
<TermTyp>Term
<TermRef>JC
**
<Created By>g04
<Subject>statistics ST, technology TE
<SourceDoc&Lang>EN/13/95
<Personal Comment>GlobeDisk BTBABR940001264
<EN>CIS
<TermTyp>Abrev.
<EN>Community Innovation Survey
<TermTyp>Term
<ES>Encuesta de la Innovación Comunitaria
<TermTyp>Term
<TermRef>aq
**
<Created By>g04
<Subject>intellect. property JUD
<SourceDoc&Lang>EN/13/50110000/95
<Personal Comment>GlobeDisk MBZABR910011164
<EN>IPR
<TermTyp>Abrev.
<EN>Intellectual Property Rights
<TermTyp>Term
<ES>derechos de propiedad industrial e intelectual
<TermTyp>Term
<TermRef>mt
<ES>DPII
<TermTyp>Abrev.
**
<Created By>g04
<Subject>technology TE
<SourceDoc&Lang>EN/13/05200000.p00/94
<Personal Comment>GlobeDisk BTLABR932000030
<EN>MINT

<TermTyp>Abrev.
<EN>Managing the Integration of New Technologies
<TermTyp>Term
<Note>SPRINT
<ES>Gestión de la integración de nuevas tecnologías
<TermTyp>Term
<TermRef>JC
<ES>MINT
<TermTyp>Abrev.
**
<Created By>g04
<Subject>informatics AUL, research and development TE4
<SourceDoc&Lang>EN/13/95/52270101.P00
<Personal Comment>GlobeDisk BTLABR930200794/ BE= BTL TY= TFI93 NI= 0020040
<EN>Multilingual Action Plans
<TermTyp>Term
<Note>Multilingüismo
<EN>MLAP
<TermTyp>Abrev.
<ES>Planes de Acción Multilingüe
<TermTyp>Term
<TermRef>JC
<ES>MLAP
<TermTyp>Abrev.
<FR>MLAP
<TermTyp>Abrev.
<FR>Plans d'Action Multilingue
<TermTyp>Term
<TermRef>FR/13/95/52270101.P00
**
<Created By>g04
<Subject>information networks OOC
<SourceDoc&Lang>EN/13/1257/94
<Personal Comment>GlobeDisk BTLABR932000165
<EN>NAP
<TermTyp>Abrev.
<EN>National Awareness Partner
<TermTyp>Term
<Note>IMPACT
<ES>colaborador nacional de difusión
<TermTyp>Term
<TermRef>JC
**
<Created By>g04
<Subject>technology TE
<SourceDoc&Lang>EN/13/95/53930000.P00

<Personal Comment>GlobeDisk TAIABR864000828
<EN>NTBF
<TermTyp>Abrev.
<EN>New Technology Based Firms
<TermTyp>Term
<Note>diseño
<ES>empresas de nuevas tecnologías
<TermTyp>Term
<TermRef>JC

B.3.1 MultiTerm to Blind MARTIF Mapping

Field	Description	Examples	12620 datCat
**	Beginning of TermEntry		<termEntry>
<Created By>	Responsibility	<Created By>g04	<ptr type='originator' target=Y/>
<SourceDoc&Lang>	Source Identifier	<SourceDoc&Lang>EN/13 /1257/94	<ref type=sourceIdentifier target=Z>...</ref> or <ptr type=sourceIdentifier target=Z/>
<Personal Comment>	Explanation	<Personal Comment>GlobeDisk BTBABR930000390	<descrip type=X>...</descrip> <note>
<EN>	Language	<EN>CST	<langSet lang='…'>
<TermTyp>		<TermTyp>Term	<termNote type=termType> … </termNote>
<Note>	Note on the term	<Note>IMPACT	<termNote type='usageNote'>...</termNote>
<TermRef>		<TermRef>JC	<ptr type='sourceIdentifier' target=Y/> Should be 'responsibility'
<Creation Date>	Date and time of record creation, using the European date convention.		<date type='creation'>...</date> Not in data sample
<Change Date>	Date and time of most recent change to record.		<date type='modification'>...</date> Not in data sample
<Changed By>	UserId of person initiating most recent change to record.		<ptr type=X target=Y/> Not in data sample

<Entry Class>	MultiTerm entry class (1-8) of record. 1 is normally default (unverified data).		Disallowed in Blind MARTIF, although it may affect the process of the export routine, i.e. the entry class may appear in the body, shared references, etc.
<Graphic>	Name of graphic file associated with entry. Not used in this file, but present in all entries.		Not in data sample
<Entry Number>	Physical record number in the MultiTerm file.		<termEntry id='…'> Not in data sample
<Project>	A project designation.		<revisionDesc> Not in data sample
<Subject>	One or more 2- or 3- letter Lennoch subject codes with short text explanation.		<ref type='subjectField' target=Y>...</ref>
<Definition>	Definition of the term		<descrip type='definition'>...</descrip>
<DefRef>	Reference for the definition		<ptr type='sourceIdentifier' target=Y/>
<Note>	Note on the term		<termNote type=X>...</termNote> <note>
<NoteRef>	Reference for the note.		<termNote type=X>...</termNote> <note>

B.4 Sample NACE (NACE) data

```
<entry>
<FR>Agriculture, chasse, sylviculture</>
<EN>Agriculture, hunting and forestry</>
<DE>Land- und Forstwirtschaft</>
<IT>Agricoltura, caccia e silvicoltura</>
<NL>Landbouw, jacht en bosbouw</>
<PT>Agricultura, produção animal, caça e silvicultura</>
<SV>Jordbruk, jakt och skogsbruk</>
<CODE>Section A</>
</entry>

<entry>
<FR>Agriculture, chasse, services annexes</>
<EN>Agriculture, hunting and related service activities</>
<DE>Landwirtschaft, gewerbliche Jagd</>
<IT>Agricoltura, caccia e relativi servizi</>
<NL>Landbouw, jacht en diensten in verband met deze activiteiten</>
<PT>Agricultura, produção animal, caça e actividades dos serviços relacionados</>
<SV>Jordbruk, jakt och service i anslutning härtill</>
<CODE>01</>
</entry>

<entry>
<FR>Culture</>
<EN>Growing of crops; market gardening; horticulture</>
<DE>Pflanzenbau</>
<IT>Coltivazioni agricole; orticoltura, floricoltura</>
<NL>Teelt van akkerbouwgewassen; teelt van groenten; tuinbouw</>
<PT>Agricultura</>
<SV>Växtodling</>
<CODE>01.1</>
<OLDCODE>011</>
</entry>

<entry>
<FR>Culture de céréales; cultures industrielles</>
<EN>Growing of cereals and other crops n.e.c.</>
<DE>Ackerbau</>
<IT>Coltivazioni di cereali e di altri seminativi n.c.a.</>
<NL>Teelt van granen en andere akkerbouwgewassen, n.e.g.</>
<PT>Cultura de cereais e outras culturas, n.e.</>
<SV>Odling av jordbruksväxter</>
<CODE>01.11</>
<OLDCODE>0111</>
```

</entry>

<entry>
<FR>Culture de légumes; horticulture; pépinières</>
<EN>Growing of vegetables, horticultural specialities and nursery products</>
<DE>Gartenbau</>
<IT>Coltivazione di ortaggi, specialità orticole, fiori e prodotti di vivai</>
<NL>Teelt van groenten, tuinbouw- en kwekerijprodukten</>
<PT>Horticultura, especialidades hortícolas e produtos de viveiro</>
<SV>Odling av köks-, prydnads- och plantskoleväxter</>
<CODE>01.12</>
<OLDCODE>0112</>
</entry>

<entry>
<FR>Culture de fruits</>
<EN>Growing of fruit, nuts, beverage and spice crops</>
<DE>Dauerkulturbau</>
<IT>Coltivazione di frutta, frutta a guscio, prodotti destinati alla preparazione di bevande e spezie</>
<NL>Teelt van fruit, noten, specerijgewassen en gewassen bestemd voor de vervaardiging van dranken</>
<PT>Culturas de frutos, de frutos de casca rija, de produtos destinados à preparação de bebidas e de especiarias</>
<SV>Odling av frukt, bär, nötter, kryddväxter m.m.</>
<CODE>01.13</>
<OLDCODE>0113</>
</entry>

<entry>
<FR>Élevage</>
<EN>Farming of animals</>
<DE>Tierhaltung</>
<IT>Allevamento di animali</>
<NL>Veeteelt</>
<PT>Produção animal</>
<SV>Djurskötsel</>
<CODE>01.2</>
<OLDCODE>012</>
</entry>

<entry>
<FR>Élevage de bovins</>
<EN>Farming of cattle, dairy farming</>
<DE>Haltung von Rindern</>
<IT>Allevamento di bovini e bufalini, produzione di latte crudo</>

```
<NL>Rundveehouderij</>
<PT>Bovinicultura</>
<SV>Nötkreatursskötsel</>
<CODE>01.21</>
<OLDCODE>0121x</>
</entry>
```

B.4.1 NACE to Blind MARTIF Mapping

Field	Description	Examples	12620 datCat
<entry>	Term Entry	<entry>	<termEntry>
<FR>	Language	<FR> French <EN> English <DE> German <IT> Italian <NL> Dutch <PT> Portuguese <SV>Swedish	<langSet lang='…'>
<CODE>	Represents the numeric code of the entry in the NACE system	<CODE>01.22</>	<termEntry id='…'>
<OLDCODE>	Represents the numeric code in the old NACE system	<OLDCODE>0121x</>	<ref type=custom target=OLDCODE> … </ref>
</>	Carriage return (end tag)		</termEntry>

Appendix C: Conversions to Blind MARTIF (export)

C.1 GlobeDisk to Blind MARTIF conversion (source code on CD-ROM)

C.1.1 Output

```
<?XML version='1.0'?>
<!DOCTYPE martif PUBLIC "ISO 12200:1997//DTD for MARTIF (Part 2 V1)//EN" [
<!ENTITY % latin1.ents PUBLIC
'ISO 8879-1986//ENTITIES Added Latin 1//EN' >
%latin1.ents;
<!ENTITY % datcats PUBLIC
'ISO 12200:1997//ENTITIES for MARTIF data categories (Part 2 V1)//EN'>
]>

<martif type=Part2v1 lang=EN>

<martifHeader>

<fileDesc>
<titleStmt><title>Glob2BTF output file</title></titleStmt>
<publicationStmt><p>converted by Glob2BTF on 1998-10-11</p></publicationStmt>
<sourceDesc><p>from file D:\AARON\Ling 699\BTF Conversions\Samples\Calvo1.txt</p></sourceDesc>
</fileDesc>

<encodingDesc>
<p type=DatCatSetName>maxi-set</p>
<p type=DatCatSetVersion>1</p>
<p type=languageDeclaration id=EN></p>
<p type=languageDeclaration id=ES></p>
</encodingDesc>
<revisionDesc>
<change><p>1998-10-11: converted from D:\AARON\Ling 699\BTF
Conversions\Samples\Calvo1.txt</p></change>
<change><p>1996-02-23: last validation of data in GlobeDisk format</p></change>
</revisionDesc>
</martifHeader>

<text>
<body>

<termEntry>
<langSet lang=EN>
 <ntig><termGrp>
  <term>European Cooperative Research Action for Technology</term>
  <termNote type='termType'>full form</termNote>
  <note>IDT</note>
 </termGrp></ntig>
 <ntig><termGrp>
  <term>CRAFT</term>
  <termNote type='termType'>acronym</termNote>
 </termGrp></ntig>
</langSet>

<langSet lang=ES>
```

```
<ntig><termGrp>
  <term>Programa europeo de investigaci½n tecnol½gica en cooperativa</term>
  <termNote type='termType'>full form</termNote>
  <ptr type=responsibility target='JC'/>
  <note>13/94. Este programa tiene por objetivo aportar  apoyo financiero a grupos de empresas industriales -
principalmente PYME-, unidas por la necesidad comœn de seguir haciendo investigaci½n industrial o
tecnol½gica, que les permita asociarse y encargar a terceros (centros de investigaci½n, universidades u otras
empresas) la ejecuci½n de un contrato de investigaci½n y desarrollo tecnol½gico (IDT) para ellas. (doc.
XXIII/94/ 05200000.p00)</note>
 </termGrp></ntig>
 <ntig><termGrp>
  <term>CRAFT</term>
  <termNote type='termType'>acronym</termNote>
 </termGrp></ntig>
</langSet>
</termEntry>

<termEntry>
<langSet lang=EN>
 <ntig><termGrp>
  <term>push-pull activity</term>
  <termNote type='termType'>full form</termNote>
  <note>VALUE</note>
 </termGrp></ntig>
</langSet>

<langSet lang=ES>
<ntig><termGrp>
  <term>actividad de empuje-tir½n</term>
  <termNote type='termType'>full form</termNote>
  <ptr type=responsibility target='JC'/>
  <note>13/14490000/94</note>
 </termGrp></ntig>
</langSet>
</termEntry>

<termEntry>
<langSet lang=EN>
 <ntig><termGrp>
  <term>R&D Limited Partnerships</term>
  <termNote type='termType'>full form</termNote>
  <note>INNOVATION</note>
 </termGrp></ntig>
</langSet>

<langSet lang=ES>
<ntig><termGrp>
  <term>Sociedades Limitadas de I+D</term>
  <termNote type='termType'>full form</termNote>
  <ptr type=responsibility target='ar'/>
  <note>13/95/53650000.P00</note>
 </termGrp></ntig>
</langSet>
</termEntry>

<termEntry>
```

```
<langSet lang=EN>
 <ntig><termGrp>
  <term>task force</term>
  <termNote type='termType'>full form</termNote>
  <note>VALUE</note>
 </termGrp></ntig>
</langSet>

<langSet lang=ES>
<ntig><termGrp>
  <term>grupo operativo</term>
  <termNote type='termType'>full form</termNote>
  <ptr type=responsibility target='JC'/>
  <note>13/14520000.p00</note>
 </termGrp></ntig>
</langSet>
</termEntry>

........ REMAINDER OF THIS CONVERSION WILL BE AVAILABLE ON CD-ROM ........

</body>

<back>

<refObjectList type=responsibleParty>
<refObject id='JC'>
<item type=name>JC</item>
</refObject>
<refObject id='ar'>

<item type=name>ar</item>
</refObject>
<refObject id='AQ'>
<item type=name>AQ</item>
</refObject>
<refObject id='candemi'>
<item type=name>candemi</item>
</refObject>
<refObject id='mt'>
<item type=name>mt</item>
</refObject>
<refObject id='MC'>
<item type=name>MC</item>
</refObject>
</refObjectList>
</back>
</text>
</martif>
```

C.2 Eurodicautom to Blind MARTIF conversion (source code on CD-ROM)

C.2.1 Output

```
<?XML version='1.0'?>
<!DOCTYPE martif PUBLIC "ISO 12200:1997//DTD for MARTIF (Part 2 V1)//EN" [
<!ENTITY % latin1.ents PUBLIC
'ISO 8879-1986//ENTITIES Added Latin 1//EN' >
%latin1.ents;
<!ENTITY % datcats PUBLIC
'ISO 12200:1997//ENTITIES for MARTIF data categories (Part 2 V1)//EN'>
]>

<martif type=Part2v1 lang=DE>

<martifHeader>

<fileDesc>
<titleStmt><title>EDT2BTF output file</title></titleStmt>
<publicationStmt><p>converted by EDT2BTF on 1998-10-11</p></publicationStmt>
<sourceDesc><p>from file D:\AARON\Ling 699\BTF Conversions\Samples\Agl03.eco</p></sourceDesc>
</fileDesc>

<encodingDesc>
<p type=DatCatSetName>maxi-set</p>
<p type=DatCatSetVersion>1</p>
<p type=languageDeclaration id=DE></p>
<p type=languageDeclaration id=FR></p>
<p type=languageDeclaration id=EN></p>
<p type=languageDeclaration id=DA></p>
<p type=languageDeclaration id=ES></p>
<p type=languageDeclaration id=EL></p>
<p type=languageDeclaration id=NL></p>
<p type=languageDeclaration id=PT></p>
<p type=languageDeclaration id=FI></p>
<p type=languageDeclaration id=SV></p>
</encodingDesc>

<revisionDesc>
<change><p>1998-10-11: converted from D:\AARON\Ling 699\BTF
Conversions\Samples\Agl03.eco</p></change>
</revisionDesc>
</martifHeader>

<text>
<body>


<termEntry id=ID3>
<date type=origination>1996-05-24T16:12:02</date>
<ptr type=originator target='agl03@hamesae'/>
<date type=modification>1996-06-18T15:05:12</date>
<ptr type=updater target='agl03@hamesae'/>
<ref type=subjectField target='sfME'>medicine</ref>
```

```
<langSet lang=DE>
 <ntig><termGrp>
  <term>Gesamt-Aerosole</term>
  <termNote type='termType'>full form</termNote>
  <termNote type=reliabilityCode>4</termNote>
  <ptr type=sourceIdentifier target='Dok.039495'/>
 </termGrp>
</ntig>
</langSet>

<langSet lang=FR>
<ptr type=sourceIdentifier target='Dok.039495'/>
 <ntig><termGrp>
  <term>concentration atmosph&#x00E9;rique totale</term>
  <termNote type='termType'>full form</termNote>
  <termNote type=reliabilityCode>4</termNote>
  <termNote type=usageNote>a&#x00E9;rosol solide</termNote>
  <ref type=sourceIdentifier target='R.Lauwerys'>p.10</ref>
 </termGrp>
</ntig>
</langSet>

</termEntry>

<termEntry id=ID4>
<date type=origination>1996-05-24T16:17:03</date>
<ptr type=originator target='agl03@hamesae'/>
<date type=modification>1996-06-18T15:06:45</date>
<ptr type=updater target='agl03@hamesae'/>
<ref type=subjectField target='sfME'>medicine</ref>
<langSet lang=DE>
 <ntig><termGrp>
  <term>Aerosol-Durchmesser</term>
  <termNote type='termType'>full form</termNote>
  <termNote type=reliabilityCode>4</termNote>
  <ptr type=sourceIdentifier target='Dok.0394950'/>
 </termGrp>
</ntig>
</langSet>

<langSet lang=FR>
<ptr type=sourceIdentifier target='Dok.039495'/>
 <ntig><termGrp>
  <term>diam&#x00E8;tre des particules de l'a&#x00E9;rosol</term>
  <termNote type='termType'>full form</termNote>
  <termNote type=reliabilityCode>4</termNote>
  <termNote type=usageNote>diam&#x00E8;tre a&#x00E9;rodynamique des particules</termNote>
  <ref type=sourceIdentifier target='R.Lauwerys'>p.16</ref>
 </termGrp>
</ntig>
</langSet>

</termEntry>

<termEntry id=ID6>
<date type=origination>1996-05-28T16:49:25</date>
```

```
<ptr type=originator target='agl03@hamesae'/>
<date type=modification>1996-06-18T15:13:06</date>
<ptr type=updater target='agl03@hamesae'/>
<ref type=subjectField target='sfAG'>agriculture</ref>
<ref type=subjectField target='sfST'>statistics</ref>
<langSet lang=EN>
 <ntig><termGrp>
  <term>objective yield model</term>
  <termNote type='termType'>full form</termNote>
  <termNote type=reliabilityCode>4</termNote>
  <ptr type=sourceIdentifier target='OS166595en'/>
 </termGrp>
</ntig>
</langSet>

<langSet lang=FR>
<ptr type=sourceIdentifier target='OS166595en'/>
 <ntig><termGrp>
  <term>mod&#x00E8;le de pr&#x00E9;vision de rendements objectifs</term>
  <termNote type='termType'>full form</termNote>
  <termNote type=reliabilityCode>4</termNote>
  <termNote type=usageNote>pr&#x00E9;vision des rendements de ma&#x00EF;s et de soja aux
USA</termNote>
  <ptr type=sourceIdentifier target='OSdemandeu'/>
 </termGrp>
</ntig>
</langSet>

</termEntry>

<termEntry id=ID7>
<date type=origination>1996-05-28T16:54:46</date>
<ptr type=originator target='agl03@hamesae'/>
<date type=modification>1996-06-18T15:02:19</date>
<ptr type=updater target='agl03@hamesae'/>
<ref type=subjectField target='sfAG'>agriculture</ref>
<ref type=subjectField target='sfST'>statistics</ref>
<langSet lang=EN>
 <ntig><termGrp>
  <term>area frame</term>
  <termNote type='termType'>full form</termNote>
  <termNote type=reliabilityCode>4</termNote>
  <ptr type=sourceIdentifier target='OS166595en'/>
 </termGrp>
</ntig>
</langSet>

<langSet lang=FR>
<ptr type=sourceIdentifier target='OS166595en'/>
 <ntig><termGrp>
  <term>base de sondage ar&#x00E9;olaire</term>
  <termNote type='termType'>full form</termNote>
  <termNote type=reliabilityCode>4</termNote>
  <termNote type=usageNote>pr&#x00E9;vision des rendements de ma&#x00EF;s et de soja aux
USA</termNote>
  <ptr type=sourceIdentifier target='OSdemandeu'/>
```

```
 </termGrp>
</ntig>
</langSet>

</termEntry>

........ REMAINDER OF THIS CONVERSION WILL BE AVAILABLE ON CD-ROM ........

</body>

<back>
<refObjectList type='responsibleParty'>
<refObject id='agl03hames'>
<item type='name'>agl03@hamesae</item>
</refObject>
<refObject id='agl03reisa'>
<item type='name'>agl03@reisant</item>
</refObject>
<refObject id='agl03benne'>
<item type='name'>agl03@bennero</item>
</refObject>
<refObject id='agl03jense'>
<item type='name'>agl03@jensein</item>
</refObject>
</refObjectList>

<refObjectList type='subjectFieldSet'>
<refObject id='sfME'>
<item type='subjectFieldName'>medicine</item>
</refObject>
<refObject id='sfAG'>
<item type='subjectFieldName'>agriculture</item>
</refObject>
<refObject id='sfST'>
<item type='subjectFieldName'>statistics</item>
</refObject>
<refObject id='sfTV'>
<item type='subjectFieldName'>labour</item>
</refObject>
<refObject id='sfJU'>
<item type='subjectFieldName'>law</item>
</refObject>
<refObject id='sfAS'>
<item type='subjectFieldName'>insurance</item>
</refObject>
<refObject id='sfUL'>
<item type='subjectFieldName'>informatics </item>
</refObject>
<refObject id='sfEN'>
<item type='subjectFieldName'>environment</item>
</refObject>
<refObject id='sfCH'>
<item type='subjectFieldName'>chemistry/chem. ind.</item>
</refObject>
<refObject id='sfCA'>
<item type='subjectFieldName'>economic developm. </item>
```

```xml
</refObject>
<refObject id='sfPO'>
<item type='subjectFieldName'>politics</item>
</refObject>
<refObject id='sfBA'>
<item type='subjectFieldName'>construction ind.</item>
</refObject>
<refObject id='sfTR'>
<item type='subjectFieldName'>transport</item>
</refObject>
<refObject id='sfMG'>
<item type='subjectFieldName'>mechanical engin.</item>
</refObject>
<refObject id='sfEC'>
<item type='subjectFieldName'>economics</item>
</refObject>
<refObject id='sfFI'>
<item type='subjectFieldName'>finance</item>
</refObject>
<refObject id='sfLC'>
<item type='subjectFieldName'>energy </item>
</refObject>
<refObject id='sfTE'>
<item type='subjectFieldName'>technology</item>
</refObject>
<refObject id='sfAD'>
<item type='subjectFieldName'>administration</item>
</refObject>
<refObject id='sfED'>
<item type='subjectFieldName'>education</item>
</refObject>
</refObjectList>

<refObjectList type='bibl'>
<refObject id='Dok.039495'>
<item type='title'>Dok. 0394/95de</item>
</refObject>
<refObject id='R.Lauwerys'>
<item type='title'>R. Lauwerys, Toxicologie industrielle</item>
</refObject>
<refObject id='Dok.0394950'>
<item type='title'>Dok. 0394/95</item>
</refObject>
<refObject id='OS166595en'>
<item type='title'>OS/1665/95en</item>
</refObject>
<refObject id='OSdemandeu'>
<item type='title'>OS/demandeur</item>
</refObject>
<refObject id='OS321195de'>
<item type='title'>OS/3211/95de</item>
</refObject>
<refObject id='CEMAGREFDi'>
<item type='title'>CEMAGREF, Dict. techn. machinisme &#x0026; &#x00E9;quipements agricoles</item>
</refObject>
<refObject id='CNEEMADict'>
```

```
<item type='title'>CNEEMA, Dict. techn. de la m&#x00E9;canisation agricole</item>
</refObject>
<refObject id='IMAGGuided'>
<item type='title'>IMAG, Guide de m&#x00E9;canisation</item>
</refObject>
<refObject id='AGL3-AHH'>
<item type='title'>AGL3-AHH</item>
</refObject>
<refObject id='Larousseag'>
<item type='title'>Larousse agricole</item>
</refObject>
<refObject id='DroitSocia'>
<item type='title'>Droit Social n 5/94-457</item>
</refObject>
<refObject id='DroitSocia0'>
<item type='title'>Droit Social n 5/94-484</item>
</refObject>
<refObject id='OPOCE00429'>
<item type='title'>OPOCE 0042/96fr</item>
</refObject>
<refObject id='OleBrinchG'>
<item type='title'>Ole Brinch, Grafisk H&#x00F8;jskole</item>
</refObject>
<refObject id='OPOCE004290'>
<item type='title'>OPOCE 0042/96 fr</item>
</refObject>
<refObject id='DroitSocia1'>
<item type='title'>Droit Social n 5/94-497</item>
</refObject>
<refObject id='DroitSocia2'>
<item type='title'>Droit Social n 5/94-502</item>
</refObject>
<refObject id='DroitSocia3'>
<item type='title'>Droit Social n 5/94-503</item>
</refObject>
<refObject id='OPOCE00419'>
<item type='title'>OPOCE 0041/96FR</item>
</refObject>
<refObject id='DroitSocia4'>
<item type='title'>Droit Social n 5/94-504</item>
</refObject>
<refObject id='DroitSocia5'>
<item type='title'>Droit Social n 5/94-506</item>
</refObject>
<refObject id='DroitSocia6'>
<item type='title'>Droit Social n 5/94-507, 508</item>
</refObject>
<refObject id='DroitSocia7'>
<item type='title'>Droit Social n 5/94-507</item>
</refObject>
<refObject id='DroitSocia8'>
<item type='title'>Droit Social n 5/94-508</item>
</refObject>
<refObject id='DroitSocia9'>
<item type='title'>Droit Social n 5/94-509</item>
</refObject>
```

```xml
<refObject id='DroitSocia10'>
<item type='title'>Droit Social n 4/94-349</item>
</refObject>
<refObject id='DroitSocia11'>
<item type='title'>Droit Social n 4/94-422</item>
</refObject>
<refObject id='DroitSocia12'>
<item type='title'>Droit Social n 4/94-368</item>
</refObject>
<refObject id='PtroleetTe'>
<item type='title'>P&#x00E9;trole et Techniques n&#x00B0;394/1995-63</item>
</refObject>
<refObject id='PtroleetTe0'>
<item type='title'>P&#x00E9;trole et Techniques n&#x00B0;394/1995-82</item>
</refObject>
<refObject id='PtroleetTe1'>
<item type='title'>P&#x00E9;trole et Techniques n&#x00B0;394/1995-84</item>
</refObject>
<refObject id='LW20-09-95'>
<item type='title'>LW 20-09-95</item>
</refObject>
<refObject id='LW13-09-95'>
<item type='title'>LW, 13-09-95</item>
</refObject>
<refObject id='Doc.055481'>
<item type='title'>Doc. 05/5481/96en</item>
</refObject>
<refObject id='cf.OMSVoca'>
<item type='title'>cf. OMS, Vocabulaire de la sant&#x00E9; publique</item>
</refObject>
<refObject id='cf.Manuila'>
<item type='title'>cf. Manuila</item>
</refObject>
<refObject id='CdJAffaire'>
<item type='title'>CdJ, Affaire C-242/95-n&#x00B0;8</item>
</refObject>
<refObject id='J.ChaponTr'>
<item type='title'>J. Chapon, Travaux Maritimes 2-134</item>
</refObject>
<refObject id='V540896fr'>
<item type='title'>V5408/96fr</item>
</refObject>
<refObject id='M.BarbierD'>
<item type='title'>M. Barbier, Dict. techniqiue du b&#x00E2;timent et des travaux publics</item>
</refObject>
<refObject id='GDEL'>
<item type='title'>GDEL</item>
</refObject>
<refObject id='M.BarbierD0'>
<item type='title'>M.Barbier, Dict. technique du b&#x00E2;timent et des travaux publics</item>
</refObject>
<refObject id='B.Dubuisso'>
<item type='title'>B. Dubuisson, Enc. pratique de la construction et du b&#x00E2;timent I-1226+1229</item>
</refObject>
<refObject id='B.Dubuisso0'>
```

```xml
<item type='title'>B. Dubuisson, Enc. pratique de la construction et du b&#x00E2;timent I - 1231
fig.44+1232</item>
</refObject>
<refObject id='Ptroleette2'>
<item type='title'>P&#x00E9;trole et techniques, n&#x00B0;399/96-42</item>
</refObject>
<refObject id='doc.095555'>
<item type='title'>doc. 09/5555/95de</item>
</refObject>
<refObject id='DFBLescloi'>
<item type='title'>DFB, Les cloisons, fiche techn. 2.42</item>
</refObject>
<refObject id='R.BayonAMD'>
<item type='title'>R. Bayon, AMD, Les cloisons ds le b&#x00E2;t</item>
</refObject>
<refObject id='DFBLescloi0'>
<item type='title'>DFB, Les cloisons</item>
</refObject>
<refObject id='DFBLescloi1'>
<item type='title'>DFB, Les cloisons, fiche techn. 7.02</item>
</refObject>
<refObject id='DFBLescloi2'>
<item type='title'>DFB, Les cloisons, fiche techn. 7.01</item>
</refObject>
<refObject id='cf.DFBLesc'>
<item type='title'>cf. DFB, Les cloisons</item>
</refObject>
<refObject id='ErnstWoert'>
<item type='title'>Ernst, Woerterb. der ind. Technik</item>
</refObject>
<refObject id='cf.Bucksch'>
<item type='title'>cf. Bucksch, Dict. pour l'Arch., le B&#x00E2;t. et les Mat. de Constr</item>
</refObject>
<refObject id='BuckschDic'>
<item type='title'>Bucksch, Dict. pour l'Arch., le B&#x00E2;t. et les Mat</item>
</refObject>
<refObject id='ErnstDict.'>
<item type='title'>Ernst, Dict. de la techn. ind</item>
</refObject>
<refObject id='DGIXLuxbg'>
<item type='title'>DG IX, Luxbg</item>
</refObject>
<refObject id='DFBLescloi3'>
<item type='title'>DFB, Les cloisons, fiche technique 7.02</item>
</refObject>
<refObject id='DFBLescloi4'>
<item type='title'>DFB, Les cloisons, fiche technique 2.42</item>
</refObject>
<refObject id='BuckschDic0'>
<item type='title'>Bucksch, Dict. pour l'Arch., le B&#x00E2;t. et les Mat. de Constr</item>
</refObject>
<refObject id='DGIXLuxemb'>
<item type='title'>DG IX, Luxembourg</item>
</refObject>
<refObject id='R.BayonLes'>
<item type='title'>R. Bayon, Les cloisons ds le b&#x00E2;t</item>
```

</refObject>
<refObject id='R.BayonLes0'>
<item type='title'>R. Bayon, Les cloisons dans le b&#x00E2;timent</item>
</refObject>
<refObject id='W.Meyer-Bo'>
<item type='title'>W.Meyer-Bohe, Innenausbau-Trennw&#x00E4;nde, S.42</item>
</refObject>
<refObject id='Euroscript'>
<item type='title'>Euroscript&#x003C;MO84949-EN&#x003E;</item>
</refObject>
<refObject id='JOL34094-'>
<item type='title'>JO L340/94-</item>
</refObject>
<refObject id='EMEA010896'>
<item type='title'>EMEA 0108/96</item>
</refObject>
<refObject id='L.Perlemut'>
<item type='title'>L.Perlemuter, Dict. pratique de pharmacologie clinique</item>
</refObject>
<refObject id='L.Perlemut0'>
<item type='title'>L. Perlemuter, Dict. pratique de pharmacologie clinique</item>
</refObject>
<refObject id='JOL21193-8'>
<item type='title'>JO L211/93-8</item>
</refObject>
<refObject id='JOL14395Rg'>
<item type='title'>JO L143/95, R&#x00E8;glement 1441/95</item>
</refObject>
<refObject id='AGL3'>
<item type='title'>AGL3</item>
</refObject>
<refObject id='Pharmainfo'>
<item type='title'>Pharmainformation, Innsbruck 9/1</item>
</refObject>
<refObject id='Doc.orig.C'>
<item type='title'>Doc. orig. CdJ</item>
</refObject>
<refObject id='Subdirecci'>
<item type='title'>Subdirecci&#x00F3;n General de Evaluaci&#x00F3;n de Medicamentos. Ministerio de Sanidad y Consumo</item>
</refObject>
<refObject id='Perlemuter'>
<item type='title'>Perlemuter, Dict. pratique de pharmacologie clinique</item>
</refObject>
<refObject id='ManuilaDic'>
<item type='title'>Manuila, Dict. fran&#x00E7;ais de m&#x00E9;decine et de biologie</item>
</refObject>
<refObject id='Vidal1991'>
<item type='title'>Vidal 1991</item>
</refObject>
<refObject id='ID0'>
<item
type='title'>&#x0395;&#x03C5;&#x03C1;&#x03C9;&#x03C0;&#x03B1;&#x03CA;&#x03BA;&#x03CC;&#x03C4;&#x03B5;&#x03BB;&#x03C9;&#x03BD;&#x03B5;&#x03B9;&#x03B1;&#x03BA;&#x03CC;&#x03B5;&#x03C5;&#x03C1;&#x03B5;&#x03C4;&#x03AE;&#x03C1;&#x03B9;&#x03BF;&#x03C7;&#x03B7;&#x03BC;&#x03B9;&#x03BA;&#x03CE;&#x03BD;

ουσιών και
εθνικό;
συνταγολόγιο</item>
</refObject>
<refObject id='Europesedo'>
<item type='title'>Europese douanelijst van chemische stoffen</item>
</refObject>
<refObject id='Dra.MariaT'>
<item type='title'>Dra. Maria Teresa Costa, Ministério da Saúde</item>
</refObject>
</refObjectList>

</back>
</text>
</martif>

C.3 MultiTerm to Blind MARTIF conversion (source code on CD-ROM)

C.3.1 Output

```
<?XML version='1.0'?>
<!DOCTYPE martif PUBLIC "ISO 12200:1997//DTD for MARTIF (Part 2 V1)//EN" [
<!ENTITY % latin1.ents PUBLIC
'ISO 8879-1986//ENTITIES Added Latin 1//EN' >
%latin1.ents;
<!ENTITY % datcats PUBLIC
'ISO 12200:1997//ENTITIES for MARTIF data categories (Part 2 V1)//EN'>
]>

<martif type=Part2v1 lang=EN>

<martifHeader>

<fileDesc>
<titleStmt><title>MME2BTF output file</title></titleStmt>
<publicationStmt><p>converted by MME2BTF on 1998-10-11</p></publicationStmt>
<sourceDesc><p>from file D:\AARON\Ling 699\BTF Conversions\Samples\G04.txt</p></sourceDesc>
</fileDesc>

<encodingDesc>
<p type=DatCatSetName>maxi-set</p>
<p type=DatCatSetVersion>1</p>
<p type=languageDeclaration id=EN></p>
<p type=languageDeclaration id=ES></p>
<p type=languageDeclaration id=FR></p>
</encodingDesc>

<revisionDesc>
<change><p>1998-10-11: converted from D:\AARON\Ling 699\BTF
Conversions\Samples\G04.txt</p></change>
</revisionDesc>
</martifHeader>

<text>
<body>


<termEntry>
<ptr type=originator target='g04'/>
<ptr type=sourceIdentifier target='EN-13-1257-94'/>
<note>GlobeDisk BTBABR930000390</note>
<langSet lang=en>
 <ntig><termGrp>
  <term>CST</term>
  <termNote type=termType>abbreviated form</termNote>
 </termGrp></ntig>
 <ntig><termGrp>
  <term>Central Support Team</term>
  <termNote type=termType>full form</termNote>
  <termNote type=usageNote>IMPACT</termNote>
 </termGrp></ntig>
```

```
</langSet>
<langSet lang=es>
 <ntig><termGrp>
  <term>equipo central de apoyo</term>
  <termNote type=termType>full form</termNote>
  <ptr type=sourceIdentifier target='JC'/>
 </termGrp></ntig>
</langSet>
</termEntry>


<termEntry>
<ptr type=originator target='g04'/>
<ref type=subjectField target='sfST'>statistics</ref>
<ref type=subjectField target='sfTE'>technology</ref>
<ptr type=sourceIdentifier target='EN-13-95'/>
<note>GlobeDisk BTBABR940001264</note>
<langSet lang=en>
 <ntig><termGrp>
  <term>CIS</term>
  <termNote type=termType>abbreviated form</termNote>
 </termGrp></ntig>
 <ntig><termGrp>
  <term>Community Innovation Survey</term>
  <termNote type=termType>full form</termNote>
 </termGrp></ntig>
</langSet>
<langSet lang=es>
 <ntig><termGrp>
  <term>Encuesta de la Innovaci&#x00F3;n Comunitaria</term>
  <termNote type=termType>full form</termNote>
  <ptr type=sourceIdentifier target='aq'/>
 </termGrp></ntig>
</langSet>
</termEntry>


<termEntry>
<ptr type=originator target='g04'/>
<ref type=subjectField target='sfJUD'>intellect. property</ref>
<ptr type=sourceIdentifier target='EN-13-50110000-95'/>
<note>GlobeDisk MBZABR910011164</note>
<langSet lang=en>
 <ntig><termGrp>
  <term>IPR</term>
  <termNote type=termType>abbreviated form</termNote>
 </termGrp></ntig>
 <ntig><termGrp>
  <term>Intellectual Property Rights</term>
  <termNote type=termType>full form</termNote>
 </termGrp></ntig>
</langSet>
<langSet lang=es>
 <ntig><termGrp>
  <term>derechos de propiedad industrial e intelectual</term>
  <termNote type=termType>full form</termNote>
```

```
  <ptr type=sourceIdentifier target='mt'/>
 </termGrp></ntig>
 <ntig><termGrp>
  <term>DPII</term>
  <termNote type=termType>abbreviated form</termNote>
 </termGrp></ntig>
</langSet>
</termEntry>


<termEntry>
<ptr type=originator target='g04'/>
<ref type=subjectField target='sfTE'>technology</ref>
<ptr type=sourceIdentifier target='EN-13-05200000.p00-9'/>
<note>GlobeDisk BTLABR932000030</note>
<langSet lang=en>
 <ntig><termGrp>
  <term>MINT</term>
  <termNote type=termType>abbreviated form</termNote>
 </termGrp></ntig>
 <ntig><termGrp>
  <term>Managing the Integration of New Technologies</term>
  <termNote type=termType>full form</termNote>
  <termNote type=usageNote>SPRINT</termNote>
 </termGrp></ntig>
</langSet>
<langSet lang=es>
 <ntig><termGrp>
  <term>Gesti&#x00F3;n de la integraci&#x00F3;n de nuevas tecnolog&#x00ED;as</term>
  <termNote type=termType>full form</termNote>
  <ptr type=sourceIdentifier target='JC'/>
 </termGrp></ntig>
 <ntig><termGrp>
  <term>MINT</term>
  <termNote type=termType>abbreviated form</termNote>
 </termGrp></ntig>
</langSet>
</termEntry>


<termEntry>
<ptr type=originator target='g04'/>
<ref type=subjectField target='sfAUL'>informatics</ref>
<ref type=subjectField target='sfTE4'>research and development</ref>
<ptr type=sourceIdentifier target='EN-13-95-52270101.P0'/>
<note>GlobeDisk BTLABR930200794/ BE= BTL TY= TFI93 NI= 0020040</note>
<langSet lang=en>
 <ntig><termGrp>
  <term>Multilingual Action Plans</term>
  <termNote type=termType>full form</termNote>
  <termNote type=usageNote>Multiling&#x00FC;ismo</termNote>
 </termGrp></ntig>
 <ntig><termGrp>
  <term>MLAP</term>
  <termNote type=termType>abbreviated form</termNote>
 </termGrp></ntig>
```

```
</langSet>
<langSet lang=es>
 <ntig><termGrp>
  <term>Planes de Acci&#x00F3;n Multiling&#x00FC;e</term>
  <termNote type=termType>full form</termNote>
  <ptr type=sourceIdentifier target='JC'/>
 </termGrp></ntig>
 <ntig><termGrp>
  <term>MLAP</term>
  <termNote type=termType>abbreviated form</termNote>
 </termGrp></ntig>
</langSet>
<langSet lang=fr>
 <ntig><termGrp>
  <term>MLAP</term>
  <termNote type=termType>abbreviated form</termNote>
 </termGrp></ntig>
 <ntig><termGrp>
  <term>Plans d'Action Multilingue</term>
  <termNote type=termType>full form</termNote>
  <ptr type=sourceIdentifier target='FR-13-95-52270101.P0'/>
 </termGrp></ntig>
</langSet>
</termEntry>

........ REMAINDER OF THIS CONVERSION IS AVAILABLE ON CD-ROM ........

</body>

<back>
<refObjectList type=customDatCatDefs>
 <refObject type=custom id=entryClass>
  <item type=originalName> Entry Class </item>
  <item type=description> MultiTerm entry class (1-8) of record. 1 is normally default (unverified data).
</item>
 </refObject>
</refObjectList>
<refObjectList type=bibl>
 <refObject id=EN-13-1257-94>
  <item type=title>EN/13/1257/94</item>
 </refObject>
 <refObject id=JC>
  <item type=title>JC</item>
 </refObject>
 <refObject id=EN-13-95>
  <item type=title>EN/13/95</item>
 </refObject>
 <refObject id=aq>
  <item type=title>aq</item>
 </refObject>
 <refObject id=EN-13-50110000-95>
  <item type=title>EN/13/50110000/95</item>
 </refObject>
 <refObject id=mt>
  <item type=title>mt</item>
 </refObject>
```

```
<refObject id=EN-13-05200000.p00-9>
 <item type=title>EN/13/05200000.p00/94</item>
</refObject>
<refObject id=EN-13-95-52270101.P0>
 <item type=title>EN/13/95/52270101.P00</item>
</refObject>
<refObject id=FR-13-95-52270101.P0>
 <item type=title>FR/13/95/52270101.P00</item>
</refObject>
<refObject id=EN-13-95-53930000.P0>
 <item type=title>EN/13/95/53930000.P00</item>
</refObject>
<refObject id=EN-13-94-16440800.P0>
 <item type=title>EN/13/94/16440800.P00</item>
</refObject>
<refObject id=EN\13>
 <item type=title>EN\13</item>
</refObject>
<refObject id=EN-13-95-53650000.P0>
 <item type=title>EN/13/95/53650000.P00</item>
</refObject>
<refObject id=ar>
 <item type=title>ar</item>
</refObject>
<refObject id=EN-13-95-50660000.P0>
 <item type=title>EN/13/95/50660000.P00</item>
</refObject>
<refObject id=candemi>
 <item type=title>candemi</item>
</refObject>
<refObject id=MC>
 <item type=title>MC</item>
</refObject>
<refObject id=EN-OS-48050000.p00-9>
 <item type=title>EN/OS/48050000.p00/94</item>
</refObject>
<refObject id=EN-OS-95>
 <item type=title>EN/OS/95</item>
</refObject>
<refObject id=mt-traduccinfrancesa>
 <item type=title>mt/traducci&#x00F3;n francesa</item>
</refObject>
<refObject id=FR-13>
 <item type=title>FR/13</item>
</refObject>
<refObject id=FR-13-95-53350100.P0>
 <item type=title>FR/13/95/53350100.P00</item>
</refObject>
<refObject id=FR-13-94-16440400.P0>
 <item type=title>FR/13/94/16440400.P00</item>
</refObject>
<refObject id=FR-23-94-05200000.p0>
 <item type=title>FR/23/94/05200000.p00</item>
</refObject>
<refObject id=FR-23>
 <item type=title>FR/23</item>
```

```
</refObject>
<refObject id=mmr>
 <item type=title>mmr</item>
</refObject>
<refObject id=Eurodicautom>
 <item type=title>Eurodicautom</item>
</refObject>
<refObject id=FR-OP>
 <item type=title>FR/OP</item>
</refObject>
<refObject id=FR-OS-94-57730003>
 <item type=title>FR/OS/94/57730003</item>
</refObject>
<refObject id=Fuentesestadsticas19>
 <item type=title>Fuentes estad&#x00ED;sticas, 19. 1996</item>
</refObject>
<refObject id=EN-OS-95-15290000.P0>
 <item type=title>EN/OS/95/15290000.P00</item>
</refObject>
<refObject id=EN-23-05200000.p00-9>
 <item type=title>EN/23/05200000.p00/94.</item>
</refObject>
<refObject id=EN-23-95-04620000.P0>
 <item type=title>EN/23/95/04620000.P00</item>
</refObject>
<refObject id=EN-23-95-04650000.P0>
 <item type=title>EN/23/95/04650000.P00</item>
</refObject>
<refObject id=EN-23-94-03550000.P0>
 <item type=title>EN/23/94/03550000.P00</item>
</refObject>
<refObject id=en\23\96\ws\03380000>
 <item type=title>en\23\96\ws\03380000.p00</item>
</refObject>
<refObject id=ID13-96-1308>
 <item type=title>13/96/1308</item>
</refObject>
<refObject id=DGXIII-E01>
 <item type=title>DG XIII/E01</item>
</refObject>
<refObject id=ES-13-96-14290000.W0>
 <item type=title>ES/13/96/14290000.W00 (EN)</item>
</refObject>
<refObject id=EurodicBTBABR9100008>
 <item type=title>Eurodic BTBABR910000861</item>
</refObject>
<refObject id=ES-13-94-13490000.P0>
 <item type=title>ES/13/94/13490000.P00 (EN)</item>
</refObject>
<refObject id=ID0>
 <item type=title>/23/97/00150000.W00 (FR)</item>
</refObject>
<refObject id=EN-OS-97-75950000.W0>
 <item type=title>EN/OS/97/75950000.W00</item>
</refObject>
<refObject id=EN\23\97\02210003.W0>
```

```
 <item type=title>EN\23\97\02210003.W00</item>
 </refObject>
 <refObject id=UNESCONombresySiglas>
  <item type=title>UNESCO, Nombres y Siglas de la COI</item>
 </refObject>
</refObjectList>
<refObjectList type=responsibleParty>
 <refObject id=g04>
  <item type=name>g04</item>
 </refObject>
 <refObject id=super>
  <item type=name>super</item>
 </refObject>
 <refObject id=g04calvojo>
  <item type=name>g04@calvojo</item>
 </refObject>
 <refObject id=g04martioa>
  <item type=name>g04@martioa</item>
 </refObject>
</refObjectList>
<refObjectList type=subjectFieldSet>
 <refObject id=sfST>
  <item type=subjectFieldName>statistics</item>
 </refObject>
 <refObject id=sfTE>
  <item type=subjectFieldName>technology</item>
 </refObject>
 <refObject id=sfJUD>
  <item type=subjectFieldName>intellect. property</item>
 </refObject>
 <refObject id=sfAUL>
  <item type=subjectFieldName>informatics</item>
 </refObject>
 <refObject id=sfTE4>
  <item type=subjectFieldName>research and development</item>
 </refObject>
 <refObject id=sfOOC>
  <item type=subjectFieldName>information networks</item>
 </refObject>
 <refObject id=sfELC>
  <item type=subjectFieldName>energy</item>
 </refObject>
 <refObject id=sfEC7>
  <item type=subjectFieldName>undertakings</item>
 </refObject>
 <refObject id=sfECA>
  <item type=subjectFieldName>regional development</item>
 </refObject>
 <refObject id=sfPO8>
  <item type=subjectFieldName>external relations</item>
 </refObject>
 <refObject id=sfIN>
  <item type=subjectFieldName>industry</item>
 </refObject>
 <refObject id=sfCO>
  <item type=subjectFieldName>trade</item>
```

```
</refObject>
<refObject id=sfPG>
 <item type=subjectFieldName>printing and publishing</item>
</refObject>
<refObject id=sfAG>
 <item type=subjectFieldName>agriculture</item>
</refObject>
<refObject id=sfFI8>
 <item type=subjectFieldName>money/currency</item>
</refObject>
<refObject id=sfCE>
 <item type=subjectFieldName>EU general</item>
</refObject>
<refObject id=sfEN>
 <item type=subjectFieldName>environment</item>
 </refObject>
</refObjectList>
<refObjectList type=subset>
 <refObject id=Minutes>
 <item type=subsetName>Minutes</item>
 </refObject>
 <refObject id=Report>
 <item type=subsetName>Report</item>
 </refObject>
</refObjectList>
</back>
</text>
</martif>
```

C.4 NACE to Blind MARTIF conversion (source code on CD-ROM)

C.4.1 Output

```
<?XML version='1.0'?>
<!DOCTYPE martif PUBLIC "ISO 12200:1997//DTD for MARTIF (Part 2 V1)//EN" [
<!ENTITY % latin1.ents PUBLIC
'ISO 8879-1986//ENTITIES Added Latin 1//EN' >
%latin1.ents;
<!ENTITY % datcats PUBLIC
'ISO 12200:1997//ENTITIES for MARTIF data categories (Part 2 V1)//EN'>
]>

<martif type=Part2v1 lang=FR>

<martifHeader>

<fileDesc>
<titleStmt><title>Nace2BTF output file</title></titleStmt>
<publicationStmt><p>converted by Nace2BTF on 1998-10-11</p></publicationStmt>
<sourceDesc><p>from file D:\AARON\Ling 699\BTF Conversions\Samples\Nace.txt</p></sourceDesc>
</fileDesc>

<encodingDesc>
<p type=DatCatSetName>maxi-set</p>
<p type=DatCatSetVersion>1</p>
<p type=languageDeclaration id=FR></p>
<p type=languageDeclaration id=EN></p>
<p type=languageDeclaration id=DE></p>
<p type=languageDeclaration id=IT></p>
<p type=languageDeclaration id=NL></p>
<p type=languageDeclaration id=PT></p>
<p type=languageDeclaration id=SV></p>
</encodingDesc>

<revisionDesc>
<change><p>1998-10-11: converted from D:\AARON\Ling 699\BTF
Conversions\Samples\Nace.txt</p></change>
</revisionDesc>
</martifHeader>

<text>
<body>

<termEntry id=SectionA>
<langSet lang=FR><ntig>
<termGrp><term>Agriculture, chasse, sylviculture</term></termGrp>
</ntig></langSet>
<langSet lang=EN><ntig>
<termGrp><term>Agriculture, hunting and forestry</term></termGrp>
</ntig></langSet>
<langSet lang=DE><ntig>
<termGrp><term>Land- und Forstwirtschaft</term></termGrp>
```

</ntig></langSet>
<langSet lang=IT><ntig>
<termGrp><term>Agricoltura, caccia e silvicoltura</term></termGrp>
</ntig></langSet>
<langSet lang=NL><ntig>
<termGrp><term>Landbouw, jacht en bosbouw</term></termGrp>
</ntig></langSet>
<langSet lang=PT><ntig>
<termGrp><term>Agricultura, produ&#x00E7;&#x00E3;o animal, ca&#x00E7;a e
silvicultura</term></termGrp>
</ntig></langSet>
<langSet lang=SV><ntig>
<termGrp><term>Jordbruk, jakt och skogsbruk</term></termGrp>
</ntig></langSet>
</termEntry>

<termEntry id=C-01>
<langSet lang=FR><ntig>
<termGrp><term>Agriculture, chasse, services annexes</term></termGrp>
</ntig></langSet>
<langSet lang=EN><ntig>
<termGrp><term>Agriculture, hunting and related service activities</term></termGrp>
</ntig></langSet>
<langSet lang=DE><ntig>
<termGrp><term>Landwirtschaft, gewerbliche Jagd</term></termGrp>
</ntig></langSet>
<langSet lang=IT><ntig>
<termGrp><term>Agricoltura, caccia e relativi servizi</term></termGrp>
</ntig></langSet>
<langSet lang=NL><ntig>
<termGrp><term>Landbouw, jacht en diensten in verband met deze activiteiten</term></termGrp>
</ntig></langSet>
<langSet lang=PT><ntig>
<termGrp><term>Agricultura, produ&#x00E7;&#x00E3;o animal, ca&#x00E7;a e actividades dos
servi&#x00E7;os relacionados</term></termGrp>
</ntig></langSet>
<langSet lang=SV><ntig>
<termGrp><term>Jordbruk, jakt och service i anslutning h&#x00E4;rtill</term></termGrp>
</ntig></langSet>
</termEntry>

<termEntry id=C-01.1>
<ref type=custom target=OLDCODE> 011</ref>
<langSet lang=FR><ntig>
<termGrp><term>Culture</term></termGrp>
</ntig></langSet>
<langSet lang=EN><ntig>
<termGrp><term>Growing of crops; market gardening; horticulture</term></termGrp>
</ntig></langSet>
<langSet lang=DE><ntig>
<termGrp><term>Pflanzenbau</term></termGrp>
</ntig></langSet>
<langSet lang=IT><ntig>
<termGrp><term>Coltivazioni agricole; orticoltura, floricoltura</term></termGrp>
</ntig></langSet>
<langSet lang=NL><ntig>

```
<termGrp><term>Teelt van akkerbouwgewassen; teelt van groenten; tuinbouw</term></termGrp>
</ntig></langSet>
<langSet lang=PT><ntig>
<termGrp><term>Agricultura</term></termGrp>
</ntig></langSet>
<langSet lang=SV><ntig>
<termGrp><term>V&#x00E4;xtodling</term></termGrp>
</ntig></langSet>
</termEntry>

........ REMAINDER OF THIS CONVERSION WILL BE AVAILABLE ON CD-ROM ........

</body>

<back>
<refObjectList type=customDatCatDefs>
 <refObject type=custom id=OLDCODE>
  <item type=originalName> OLDCODE </item>
  <item type=description> Old NACE system code </item>
 </refObject>
</refObjectList>
</back>
</text>
</martif>
```

Appendix D: Blind MARTIF validation program

As has been discussed in this paper, Blind MARTIF is SGML-based. While SGML has a mechanism for determining the well-formedness of a document, it provides no means to verify complete validity of a document, as defined by the *datCatDefinition* refObjectList mentioned earlier.

The Blind MARTIF Validator is a Microsoft Visual Basic 5.0 application. There are two main parts to the validator: the Initialize subroutine, and the *ValidateEntry* function. The Initialize subroutine takes four arguments: a string containing the MARTIF document header, a string containing the document's back matter, a string array of all the ID values in the file (to check cross-references), and a string containing the MARTIF type information (Part2v1, for example).

When these four items are passed to the Initialize procedure, it parses the Header and Back strings into their respective SGML trees, and searches for specific information (like the encoding description, user subset data category (datCat) description, etc.). If this information is not found, an error is generated. If it is successfully found, the two trees are stored in properties of the control, to check against later. Also, the datCat Set definition is found, either explicitly in the document's <back> matter, or in the program working directory. This, too, is stored.

The *ValidateEntry* function takes in a single termEntry as a string and parses it into an SGML tree. It then walks through the tree, element by element, and calls *ValidateElement* on each. This recursive function will then validate the element and each of its 'children,' or those elements contained within it. These it checks against several pieces of information, namely the level of occurrence, content, Target attribute (for ref and ptr elements), against

those prescribed in its datCat definition in the *datCatSet* information. Additionally, each

element is examined for a Lang, or language code. If one is present, it is checked against a

collection of valid Lang codes, as defined by the Header's *encodingDescription*. If any of

these aspects of an entry does not correspond to its definition in the *datCatSet* definition, an

error is logged in the control-public string *sErrLog*, and one is added to the error count,

which is returned by the *ValidateEntry* function when the entire entry has been checked. The

user may then, if errors are indicated, access the public *sErrLog* property to determine the

problems with the entry.

Appendix E: Test Results of Conversions to Blind MARTIF

GlobeDisk to BTF

Of the four externally supplied data samples, GlobeDisk was the most difficult to analyze. Data categories within this sample include inconsistencies in the format that present obstacles in both mapping the data categories to Blind MARTIF categories by a human, but also coding the conversion routines to be able to handle the inconsistencies in the most efficient and correct manner. For example, the GlobeDisk data categories represented in #5 below (source identifier) offered many inconsistencies, including the order in which the various numbers within the data category were presented. Naturally, this presents a problem to the human trying to map the data categories, but more importantly to the coder of the conversion routines, increasing the complexity of the software programs. To solve this particular difficulty, without having the complete specification of this format, the coder was forced to take content of this data category and treat it as a whole. For future versions of this conversion, specifications of this format will necessarily be analyzed to ensure the correct manipulation of varying data categories.

Explanation of GlobeDisk format provided by Bennett:

```
File exported from GlobeDisk database.  Bilingual English-Spanish.  Character coding ANSI.  First
line ##13ENES## indicates that this is EN-ES terminology for the Commission's DG13 (Information
Market).

Second line #ÚLTIMA VALIDACIÓN# 23.02.96 indicates date on which data was last checked.

From there on:

1 line per entry, i.e. each carriage return indicates a new entry.

Source and target are separated by #.  All text outside square brackets on the source side is the
"term", followed usually but not invariably by a note in square brackets (typically indicating
context).  Each entry has only one source term.  Target side term usually commences with a note in
square brackets (typically indicating context).  This is followed by the term itself then, usually
but not invariably, in square brackets the initials of the translator originating the term.  Each
entry may have more than one target-language term.  Target-language terms are separated by ; space
and the structure [note] term [originator] may be repeated for each term.
```

	Source Tag	Description	Example	12620 Datcat	Preserved
1	##…#	EN-ES terminology for the Commision's DG 13	##13ENES#	<langSet lang='…'> <p type=languageDeclaration ID='…'>	Yes
2	#ULTIMA VALIDACI…N #DD.MM.YY	Last date of validation	#ULTIMA VALIDACI…N #23.02.96	<revisionDesc> Perhaps attach this date to each concept entry.	Yes
3	#	Separates the source from the target		N/A – Conversion algorithm would identify all text outside of square brackets on the left side of the # as the term in English (source). All text outside of square brackets on the right side is the term in Spanish (target).	N/A
4	[…] (left side)	Typically indicates context (optional)	[VALUE]	<note> This is a difficult one as well. It is my opinion that there is no structure to this category—any text is permissible inside the square brackets.	Yes
5	[…] (right side, immediately following #)	Source (many inconsistencies in sample data)	[12/95/53650 000.P00]	<note> Left as <note> as a result of the inconsistencies within this field between entries in this data sample.	Yes
6	[…] (right side, last square brackets of target-language term)	Typically originator		<ptr type='responsibility' target='…'> <ptr type='originator' target='…'>	Yes
7	;(space)	Separates the target-language terms		This separates equal terms. Each term is treated as a synonym.	

Eurodicautom to BTF

This sample format provided only minimal difficulty in the conversion to Blind MARTIF. Two data categories from the sample were not preserved in the conversion, specifically (1) *%%BE* and (2) *%%TY*. These two data categories may be preserved in future versions of Eurodicautom to Blind MARTIF conversion, predictably as *custom* data category, probably as specific data categories that are particular to any number of data category subsets of the maximal set.

Explanation of Eurodicautom format provided by Bennett:

```
File exported from MultiTerm and converted to Eurodicautom structure (new style).
Character coding is ANSI Windows, except for Greek which is ISO 8859/7.

All entries commence %%BE.  The entry-level attributes are as follows: %%BE precedes
indication of the originating body (in this case BTL - Bureau de Terminologie,
Luxembourg - indicating the Luxembourg component of the Commission's Terminology
Unit).

%%TY is followed by a "collection"code, in this case MTW96 indicating MultiTerm for
Windows.

%%NI is followed by an entry number (not necessarily starting with 1 and potentially
with gaps in the sequence).

%%CF is followed by a number from 1 to 5 indicating a reliability level for the data
(1 lowest, 5 highest).

%%AU is followed by an abbreviated version of the name of the originating
        terminologist.

%%CM is followed by one or more two- or three-letter Lennoch subject codes.
```

Note that certain fields such as AU or CM are not strictly compulsory, but none of the entry-level fields can appear more than once per record.

The entry-level attributes are followed by one or more language blocks. Each such block begins with a language code followed by carriage return:

%%DA   Danish
%%NL   Dutch
%%EN   English
%%FI   Finnish
%%FR   French
%%DE   German
%%EL   Greek
%%IT   Italian
%%PT   Portuguese
%%ES   Spanish
%%SV   Swedish

Each language may appear only once per entry.

A language block may contain any or all of the following fields (but at least %%VE or %%AB):

%%VE space, followed by one or more "terms". Each "term"is followed by an identifying number in square brackets to permit unambiguous attachment of related fields (references, etc.). If more than one term is present, separation takes the form of ; space. Note that the first "term" in a language block may nevertheless bear the number [2] if an AB (see below) is also present.

%%AB space, followed by one or more abbreviations or acronyms. As with VE, each acronym is followed by an identifying number in square brackets and ; space serves as separator.

%%RF space, followed by one or more references. An identifier in square brackets follows each reference, ensuring that it can be matched to the relevant term, definition or note. These identifiers take the form [VE1] for terms (and abbreviations) and [DF1] for definitions and [NT1] for notes. As elsewhere ; space acts as separator.

%%DF space, followed by one or more definitions. An identifier of the form [VE1] follows each definition and ; space acts as separator.

%%NT space, followed by one or more notes. An identifier of the form [VE1] follows each conventional note and ; space acts as separator. In addition, notes in one language block in each entry (usually FR) include special "entry management" information derived from MultiTerm in "squiggly" brackets. In this case, each conventional note is preceded by {NTE} space.

The significance of the other special codes is as follows:

{CRD} space   Entry creation date and time (coded using European date conventions).

{CRU} space   The unique UserId of the terminologist creating the   entry.

{CHD) space   The date and time of the most recent change to the entry.

{CHU} space   The unique UserId of the terminologist originating the most recent change to the entry.

{ECL} space   The MultiTerm entry class of the relevant record (between 1 and 8), 1 being the lowest (default) class.

{ENR} space   The MultiTerm entry number of the relevant record. Normally, this should be the same as the NI (see above).

```
{DOM} space     Subject codes, accompanied by a short text description of the
significance of the code(s).

{DOC} space     Source document as indicated (as an entry-level attribute) in the
MultiTerm entry.

{TGT} space     Target language as indicated in MultiTerm.

A %% field preceded by * is a Greek field to be read as 8859/7 character coding.

Each %% field may only occur once within a given language block, and the only { }
code which may occur more than once in a single %% NT field (or indeed an entry) is
{NTE} space.  Carriage returns are not permitted within fields, and those at end-of-
field are terminators and should not appear in converted data.
```

	Source Tag	Description	Example	12620 Datcat	Preserved
1	%%BE	Indicates originating body	%%BE BTL	<ptr type='originatingInstitution' target='…'>	No
2	%%TY	collection code	%%TY MTW96	<ptr type=X target=Y>	No
3	%%NI	Entry Number	%%NI 3	<termEntry id='…'>	Yes
4	%%CF	Reliability code	%%CF 2	<termNote type='reliabilityCode'>	Yes
5	%%AU	Originating terminologist **	%%AU hames	<ptr type='originator' target=Y/>	Yes
6	%%CM	One or more two or three letter Lennoch subject code **	%%CM ME	<ptr type='subjectField target='…'>	Yes
7	%%DE	Language code	%%DE	<langSet lang='…'>	Yes
8	%%VE	Term	%%VE Gesamt-Aerosole [1]	<term>	Yes
9	%%RF	Reference	%%RF Dok. 0394/95de [VE1]	<ref type='sourceIdentifier' target='…'>	Yes
10	%%NT {NTE}	Note	%%NT {NTE} prévision des	<termNote type=termType>TransferComment</termNote>	Yes
11	%%AB	Abbreviation or acronym	%%AB RDT [1]	<termNote type='termType'>abbreviation</termN	Yes

				ote>	
12	{CRD}	Entry creation date and time.		<date type='origination'>...</date>	Yes
13	{CRU}	The unique UserId of the terminologist creating the entry.		<ptr type='originator' target=Y/>	Yes
14	{CHD)	The date and time of the most recent change to the entry.		<date type='modification'>...</date>	Yes
15	{CHU}	The unique UserId of the terminologist originating the most recent change to the entry.		<ptr type='updater' target=Y/>	Yes
16	{ECL}	MultiTerm entry class of the relevant record 1 <> 8), 1 being the lowest (default) class.		Disallowed in Blind MARTIF, although it may affect the process of the export routine, i.e. the entry class may appear in the body, shared references, etc.	No
17	{ENR}	MultiTerm entry number of the relevant record. Normally, this should be the same as the NI.		<termEntry id='…'>	Yes
18	{DOM}	Subject codes, accompanied by a short text description of the code(s).		<ptr type='subjectField' target='…'> Already used above	Yes
19	{DOC}	Source document as indicated (as an entry-level attribute) in the MultiTerm entry.		<ref type=sourceIdentifier target=Z>...</ref> or <ptr type=sourceIdentifier target=Z/>	Yes
20	{TGT}	Target language as indicated in MultiTerm.		<langSet lang='…'>	Yes

MultiTerm to BTF

This conversion (MultiTerm to BTF) was direct.   Some data categories from this format could have been included in the sample provided by Bennett, but were not (e.g., <Creation Date>, <Change Date>, etc.). The experiment may have represented a better analysis of Blind MARTIF as an interchange format had these data categories had been included in the sample format. However, it is certain that any coder of a MultiTerm to Blind MARTIF conversion routine would have little difficulty handling these data categories.

Explanation of MultiTerm format provided by Bennett:

```
A MultiTerm backup structure file (with underlying data structure similar to (2)).  Character coding
is ANSI.  Each entry begins with ** carriage return.  Formally, carriage returns may occur within
fields, but all field-end carriage returns are terminators only and should be stripped.

Entry-level attributes follow:

<Creation Date>   Date and time of record creation, using the European date convention.

<Created By>      UserId of creator.

<Change Date>     Date and time of most recent change to record.

<Changed By>      UserId of person initiating most recent change to record.

<Entry Class>     MultiTerm entry class (1-8) of record.  1 is normally default (unverified data).

<Graphic>         Name of graphic file associated with entry.  Not used in this file, but present in
            all entries.

<Entry Number>    Physical record number in the MultiTerm file.

<Project>         A project designation.

<Subject>         One or more 2- or 3- letter Lennoch subject codes with short text explanation.
```

```
<Source Doc & Lang>     Indication of the source document (using Commission Document numbering
conventions) and, normally, a 2-lettter code in brackets indicating the document's language.


All the above fields are present once (and only once) in each entry.


A <Personal Comment> field containing informal remarks may also be present (not more than once) but
is not required.


The term data follows in blocks.  This particular file contains only <EN> (English) and <ES>
(Spanish) data.


Each entry may contain more than one term in any language, with each term occupying a separate block
(term, attributes, definitions, notes, references, etc.).


In addition to the term itself:


<EN> ...
each block must contain (once only) the attribute <Term Typ>, which has only three possible values:
Term, Phrase or Abrev. and may contain (once only):


<TermRef> Reference for the term
<Definition> Definition of the term
<DefRef> Reference for the definition
<Note> Note on the term
<NoteRef> Reference for the note.


Cross-references may occur in <Note> fields.  Each cross-reference starts and finishes with ∧.
```

	Source Tag	Description	Example	12620 Datcat	Preserved
1	**	Beginning of TermEntry		\<termEntry\>	Yes
2	\<Created By\>	Responsibility	\<Created By\>g04	\<ptr type='originator' target=Y/\>	Yes
3	\<SourceDoc&Lang\>	Souce Identifier	\<SourceDoc&Lang\>EN/13/1257/ 94	\<ref type=sourceIdentifier target=Z\>...\</ref\> or \<ptr type=sourceIdentifier target=Z/\>	Yes
4	\<Personal Comment\>	Explanation	\<Personal Comment\>GlobeDisk BTBABR930000390	\<descrip type=X\>...\</descrip\> \<note\>	Yes
5	\<EN\>	Language	\<EN\>CST	\<langSet lang='…'\>	Yes

6	<TermTyp>	Term, Phrase, Abrev.	<TermTyp>Term	<termNote type=termType> … </termNote>	Yes
7	<Note>	Note on the term	<Note>IMPACT	<termNote type='usageNote'>...</termNote>	Yes
8	<TermRef>	Souce Identifier	<TermRef>JC	<ptr type='sourceIdentifier' target=Y/> Should be 'responsibility'	Yes
9	<Creation Date>	Date and time of record creation, using the European date convention.		<date type='creation'>...</date>	Not in sample data
10	<Change Date>	Date and time of most recent change to record.		<date type='modification'>...</date>	Not in sample data
11	<Changed By>	UserId of person initiating most recent change to record.		<ptr type=X target=Y/>	Not in sample data
12	<Entry Class>	MultiTerm entry class (1-8) of record. 1 is normally default (unverified data).		Disallowed in Blind MARTIF, although it may affect the process of the export routine, i.e. the entry class may appear in the body, shared references, etc.	N/A
13	<Graphic>	Name of graphic file associated with entry. Not used in this file, but present in all entries.			Not in sample data
14	<Entry Number>	Physical record number in the MultiTerm file.		<termEntry id='…'>	Not in sample data
15	<Project>	A project designation.		<revisionDesc>	Not in sample data
16	<Subject>	One or more 2- or 3- letter Lennoch subject codes with short text explanation.		<ref type='subjectField' target=Y>...</ref>	Yes
17	<Definition>	Definition of the term		<descrip type='definition'>...</descrip>	Yes
18	<DefRef>	Reference for the definition		<ptr type='sourceIdentifier' target=Y/>	Yes
19	<Note>	Note on the term		<termNote type=X>...</termNote> <note>	Yes
20	<NoteRef>	Reference for the note.		<termNote type=X>...</termNote> <note>	Yes

116

NACE to BTF

This sample format is the simplest of the four sample formats provided by Bennett. There were not difficulties encountered in coding the conversion routine.

Explanation of NACE format provided by Bennett:

```
NACE.TXT.  A simple, but fairly large 7-language nomenclature file in ANSI character
coding with SGML-style tags.  Each entry starts <entry> carriage return and ends
</entry> carriage return (plus an empty line).


Each language starts with the two-letter language code in angle brackets:


<FR>      French
<EN>      English
<DE>      German
<IT>      Italian
<NL>      Dutch
<PT>      Portuguese
<SV>      Swedish


and finishes </> carriage return.


In addition, each entry has a <CODE> field representing the numeric code of the
entry in the NACE system and may have an <OLDCODE> field representing the numeric
code in the old NACE system if different.
```

	Source Tag	Description	Example	12620 Datcat	Preserved
1	<entry>	Term Entry		<termEntry>	Yes
2	<FR>	Language	<FR>	<langSet lang='…'>	Yes
3	<CODE>	Represents the numeric code of the entry	<CODE>01.22</>	<termEntry id='…'>	Yes

		in the NACE system			
4	<OLDCODE>	Represents the numeric code in the old NACE system	<OLDCODE>0121x</>	<ref type=custom target=OLDCODE> … </ref>	Yes
5	</>	Carriage return (end tag)		</termEntry>	Yes

Name of Sample Data Base	Abbreviation
Blind MARTIF	BTF
GlobeDisk	GBD
Eurodicautom	ECO
MultiTerm	MME or MMI
NACE	NACE

The following are the results of the conversion from the resulting SOURCE to Blind MARTIF to SOURCE (i.e., back into the sample data formats).

GlobeDisk

From GBD to BTF to GBD**:** All information preserved.

From GBD to BTF to ECO**:** The following information is lost in this conversion:

Data Category	Reason for Loss of Information
#ULTIMA VALIDACI…N #DD.MM.YY	Should have been preserved. Programming error.
[…](left side)	Inconsistencies make difficult to determine
[…]right side, immediately following #	Inconsistencies make difficult to determine
[…]right side, last square brackets of target-language term	Inconsistencies make difficult to determine

From GBD to BTF to MMI**:** The following information is lost in this conversion:

Data Category	Reason for Loss of Information
#ULTIMA VALIDACI…N #DD.MM.YY	Should have been preserved. Programming error.

From GBD to BTF to NACE**:** The following information is lost in this conversion:

Data Category	Reason for Loss of Information
##…#	No equivalent datcat in the target data base format
#ULTIMA VALIDACI…N #DD.MM.YY	No equivalent datcat in the target data base format
[…]right side, immediately following #	No equivalent datcat in the target data base format
[…]right side, last square brackets of target-language term	No equivalent datcat in the target data base format

Eurodicautom

From ECO to BTF to ECO: The following information is lost in this conversion:

Data Category	Reason for Loss of Information
%%AB	No reason for loss
{ECL}	Disallowed in Blind MARTIF, although it may affect the process of the export routine, i.e. the entry class may appear in the body, shared references, etc.
{ENR}	This data category is repetative, i.e., it is the same as %%NI

From ECO to BTF to GBD: The following information is lost in this conversion:

Data Category	Reason for Loss of Information
%%TY	No equivalent datcat in the target data base format
%%NI	No equivalent datcat in the target data base format
%%CF	No equivalent datcat in the target data base format
%%DE	No equivalent datcat in the target data base format
%%RF	No equivalent datcat in the target data base format
%%NT {NTE}	Should have been preserved. Programming error.
%%AB	Should have been preserved. Programming error.
{CRD}	No equivalent datcat in the target data base format
{CRU}	No equivalent datcat in the target data base format
{CHD)	No equivalent datcat in the target data base format
{CHU}	No equivalent datcat in the target data base format
{ECL}	Disallowed in Blind MARTIF, although it may affect the process of the export routine, i.e. the entry class may appear in the body, shared references, etc.
{ENR}	No equivalent datcat in the target data base format
{DOM}	Inconsistencies in sample data base format make difficult to determine the content of this data category
{DOC}	No equivalent datcat in the target data base format
{TGT}	No equivalent datcat in the target data base format

From ECO to BTF to MMI: The following information is lost in this conversion:

Data Category	Reason for Loss of Information
%%BE	No equivalent datcat in the target data base format
%%TY	No equivalent datcat in the target data base format
%%CF	No equivalent datcat in the target data base format
%%AB	Should have been preserved. Programming error.
{CRU}	No equivalent datcat in the target data base format
{CHU}	No equivalent datcat in the target data base format
{ECL}	Disallowed in Blind MARTIF, although it may affect the process of the export routine, i.e. the

	entry class may appear in the body, shared references, etc.
{DOC}	Should have been preserved. Programming error.

From ECO to BTF to NACE**:** The following information is lost in this conversion:

Data Category	Reason for Loss of Information
%%BE	No equivalent datcat in the target data base format
%%TY	No equivalent datcat in the target data base format
%%CF	No equivalent datcat in the target data base format
%%AU	No equivalent datcat in the target data base format
%%CM	No equivalent datcat in the target data base format
%%RF	No equivalent datcat in the target data base format
%%NT {NTE}	No equivalent datcat in the target data base format
%%AB	No equivalent datcat in the target data base format
{CRD}	No equivalent datcat in the target data base format
{CRU}	No equivalent datcat in the target data base format
{CHD)	No equivalent datcat in the target data base format
{CHU}	No equivalent datcat in the target data base format
{ECL}	Disallowed in Blind MARTIF, although it may affect the process of the export routine, i.e. the entry class may appear in the body, shared references, etc.
{ENR}	No equivalent datcat in the target data base format
{DOM}	No equivalent datcat in the target data base format
{DOC}	No equivalent datcat in the target data base format
{TGT}	No equivalent datcat in the target data base format

MultiTerm

From MME to BTF to MMI**:** All information preserved.

From MME to BTF to GBD**:** The following information is lost in this conversion:

Data Category	Reason for Loss of Information
<Personal Comment>	Should have been preserved. Programming error.
<TermTyp>	No equivalent datcat in the target data base format
<Note>	Should have been preserved. Programming error.
<TermRef>	Should have been preserved. Programming error.
<Creation Date>	No equivalent datcat in the target data base format
<Change Date>	No equivalent datcat in the target data base format
<Changed By>	Could be preserved
<Entry Class>	No equivalent datcat in the target data base format
<Graphic>	No equivalent datcat in the target data base format
<Entry Number>	No equivalent datcat in the target data base format
<Project>	No equivalent datcat in the target data base format

\<Subject>	Inconsistencies make preservation difficult
\<Definition>	Inconsistencies make preservation difficult
\<DefRef>	No equivalent datcat in the target data base format
\<Note>	Should have been preserved. Programming error.
\<NoteRef>	Should have been preserved. Programming error.

From MME to BTF to ECO**:** The following information is lost in this conversion:

Data Category	Reason for Loss of Information
\<SourceDoc&Lang>	Should have been preserved. Programming error.
\<Personal Comment>	Should have been preserved. Programming error.
\<TermTyp>	No equivalent datcat in the target data base format
\<TermRef>	No equivalent datcat in the target data base format
\<Definition>	Should have been preserved. Programming error.
\<DefRef>	No equivalent datcat in the target data base format
\<NoteRef>	No equivalent datcat in the target data base format

From MME to BTF to NACE**:** The following information is lost in this conversion:

Data Category	Reason for Loss of Information
\<Created By>	No equivalent datcat in the target data base format
\<SourceDoc&Lang>	No equivalent datcat in the target data base format
\<Personal Comment>	No equivalent datcat in the target data base format
\<TermTyp>	No equivalent datcat in the target data base format
\<Note>	No equivalent datcat in the target data base format
\<TermRef>	No equivalent datcat in the target data base format
\<Creation Date>	No equivalent datcat in the target data base format
\<Change Date>	No equivalent datcat in the target data base format
\<Changed By>	No equivalent datcat in the target data base format
\<Entry Class>	No equivalent datcat in the target data base format
\<Graphic>	No equivalent datcat in the target data base format
\<Project>	No equivalent datcat in the target data base format
\<Subject>	No equivalent datcat in the target data base format
\<Definition>	No equivalent datcat in the target data base format
\<DefRef>	No equivalent datcat in the target data base format
\<Note>	No equivalent datcat in the target data base format
\<NoteRef>	No equivalent datcat in the target data base format

NACE

From NACE to BTF to NACE**:** All information is preserved.

From NACE to BTF to ECO**:** All information is preserved, although there are many new datcats in the ECO.

From NACE to BTF to GBD: The following information is lost in this conversion:

Data Category	Reason for Loss of Information
<FR>	No equivalent datcat in the target data base format
<CODE>	No equivalent datcat in the target data base format
<OLDCODE>	No equivalent datcat in the target data base format

From NACE to BTF to MMI: <OLDCODE> is not preserved. This is because no equivalent is found in the MMI.

Appendix F: Conversions from Blind MARTIF to sample data formats (import)

For purposes of limiting the length of this appendix, each output example is limited to three pages or less. Any additional entries not seen here are available on the CD-ROM. Also, to limit the length of this appendix, only sample conversions from GlobeDisk to Blind MARTIF to target database are shown.

F.1 Blind MARTIF to GlobeDisk conversion (source code on CD-ROM)

F.1.1 Output

##37ENES#
#ÙLTIMA VALIDACIÓN #23.02.96
European Cooperative Research Action for Technology (CRAFT) [IDT]#[13/94. Este programa tiene por objetivo aportar apoyo financiero a grupos de empresas industriales -principalmente PYME-, unidas por la necesidad comœn de seguir haciendo investigaci½n industrial o tecnol½gica, que les permita asociarse y encargar a terceros (centros de investigaci½n, universidades u otras empresas) la ejecuci½n de un contrato de investigaci½n y desarrollo tecnol½gico (IDT) para ellas. (doc. XXIII/94/ 05200000.p00)] Programa europeo de investigaci½n tecnol½gica en cooperativa (CRAFT) [JC]
push-pull activity [VALUE]#[13/14490000/94] actividad de empuje-tir½n [JC]
R&D Limited Partnerships [INNOVATION]#[13/95/53650000.P00] Sociedades Limitadas de I+D [ar]
task force [VALUE]#[13/14520000.p00] grupo operativo [JC]
awareness [IMPACT]#[13/] concienciaci½n [AQ]; [1257/94. Promotion of awareness in the EEA. (Anexo IV)] conocimiento [JC]; [Sensibilizaci½n en cuestiones medioambientales.] sensibilizaci½n [AQ]
back-up initiative [VALUE]#[13/14490000/94] iniciativa de reserva [JC]
benchmarking [full form]#[13/95/50660000.P00] patrones de referencia (establecimiento de) [candemi]
bottom-up approach [VALUE]#[13/14490000/94] enfoque ascendente [JC]
business plan [VALUE]#[13/14490000/94] plan de negocio [JC]
Central Support Team (CST) [IMPACT]#[13/1257/94. Anexo I] Equipo central de apoyo (CST) [JC]
Community Innovation Survey (CIS) [full form]#[13/] Encuesta de la Innovaci½n Comunitaria [aq]
context driven application [VALUE]#[13/14490000/94] aplicaci½n por contexto [JC]
demand pull [VALUE]#[13/14490000/94. Una innovaci½n o invento se desarrolla en funci½n de las exigencias de la demanda (concepto contrario al de Technology Push)] tir½n de la demanda [JC]
dissemination [VALUE]#[13/14490000.p00. Explotaci½n y difusi½n de los resultados de investigaci½n.] difusi½n [JC]
distribution network [VALUE]#[13/14490000/94] red de distribuci½n [JC]
European Association for the Transfer of Technology, Innovation and Industrial Information (TII) [SPRINT]#[13/Puede aparecer la traducci½n de la sigla TII como TECNOLOG™A, INNOVACI…N, INFORMACI…N] Asociaci½n Europea de Transferencia de Tecnolog-a, Innovaci½n e Informaci½n Industrial (TII) [JC]
European Commission Host Organization (ECHO) [IMPACT]#[13] Organizaci½n de distribuci½n de la Comisi½n Europea (ECHO) [JC]
European Innovation Monitoring System (EIMS) [full form]#[13/94/16440400.P00] Observatorio Europeo de la Innovaci½n (EIMS) [mt]
European Securities Dealers Association [INNOVATION]#[13/95/53650000.P00] Asociaci½n Europea de Intermediarios Financieros [ar]
event [full form]#[13/1257\94. Anexo I (seminars, workshops, exhibitions, conferences..)] acto [JC]
event schedule [contratros]#[13/1257/94] programa de actividades [JC]
ex ante [VALUE]#[13/14490000/94: integraci½n previa o ex ante de las perspectivas empresariales en los programas de IDT] ex ante [JC]
exploitation [VALUE]#[13/14490000/94. Explotaci½n y difusi½n de los resultados de investigaci½n.] explotaci½n [JC]
hardware [inform tica]#[13] equipo (s)
Help Desk [VALUE]#[13/14490000/94] Servicio de asistencia t_cnica [JC]

host [VALUE]#[13/14490000/94.Organismo, entidad, etc., como una librer-a, que acoge un servicio (por ejemplo, los centros de enlace VALUE)] organismo anfitri½n [JC]

host [inform tica]#[13/] distribuidor [JC]

Innovation Relay Centres (IRC) [difusi½n de resultados]#[13/94/17870100.P00. Centros asesores de orientaci½n pr ctica que facilitan a sus clientes el acceso a la investigaci½n y la tecnolog-a. Su objetivo principal es promocionar la transferencia de resultados de la investigaci½n y de tecnolog-as de conformidad con las necesidades manifestadas por el tejido industrial local, con objeto de mejorar su competitividad a trav_s de la innovaci½n.] centros de enlace de la innovaci½n (IRC) [JC]

Intellectual Property Rights (IPR) [full form]#[13/50110000/95] derechos de propiedad industrial e intelectual (DPII) [mt]

learning society [full form]#[textos referentes a la sociedad de la informaci½n (Libro Blanco EnseÏar y aprender: hacia la sociedad cognitiva, doc. nõ XXII/95/0159)] sociedad cognitiva [candemi]

Legal Advisory Board (LAB) [IMPACT]#[13] Grupo Consultivo Jur-dico (GCJ) [JC]

licensing [VALUE]#[13/14490000/94] concesi½n de licencias [JC]

Managing the Integration of New Technologies (MINT) [SPRINT]#[13/05200000.p00/94] Gesti½n de la integraci½n de nuevas tecnolog-as (MINT) [JC]

Multilingual Action Plans (MLAP) [Multiling_ismo]#[13/95/52270101.P00. Ces cinq plans, introduits _ partir de 1976 conjointement par la DG XIII et le SdT, ont _t_ lanc_s pour am_liorer le transfert de l'information entre langues europ_ennes, notamment par la mise en oeuvre de produits utilisant les technologies de l'information.Le sixi_me sera soumis prochainement au SdT pour approbation. (Eurodicautom). V_ase 13fres Plans d'Action Multilingue.] Planes de Acci½n Multiling_e (MLAP) [JC]

Multimedia [IMPACT]#[13/13590000.p00/94. Se utilizar el sustantivo (los) multimedios para designar el concepto. Asimismo multimedios se utilizar en funci½n adjetiva (un ordenador multimedios)] multimedios [JC]

National Awareness Partner (NAP) [IMPACT]#[13/1257/94] colaborador nacional de difusi½n (NAP) [JC]

New Technology Based Firms (NTBF) [diseÏo]#[13/95/53930000.P00] empresas de nuevas tecnolog-as [JC]

Organization for the Promotion of Energy Technologies (OPET) [Difusi½n de resultados]#[13/94/16440800.P00. La red OPET es uno de los componentes del programa THERMIE. En la actualidad la forman 49 organizaciones y se dedica a la difusi½n y el aprovechamiento de los resultados de la IDT y la demostraci½n en el campo de la energ-a, as- como a la difusi½n de tecnolog-as energ_ticas entre las empresas (sectores objetivo: energ-a, medio ambiente, transportes, industrias de la transformaci½n).] Organizaci½n de fomento de las tecnolog-as energ_ticas (OPET) [JC]

patent claim [VALUE]#[13/14490000.p00] reivindicaci½n de patente [JC]

patent filing [VALUE]#[13/14490000/94] registro de la patente [JC]

patent request [VALUE]#[13] solicitud de patente [JC]

patent specifications [VALUE]#[13/14490000.p00] descripci½n de la patente [JC]

people network [Innovaci½n]#[13/95/56170000.P00. Por redes de profesionales entendemos cualquier tipo de relaci½n informativa, formal o informal, que ponga en contacto a los miembros de una profesi½n determinada... El Programa de Innovaci½n confiere a dichos contactos dimensi½n europea...] red de profesionales [JC]

proactive [full form]#[13/Actividades de sensibilizaci½n en materia de innovaci½n y transferencia de tecnolog-a. Documentos sobre centros de enlace. DG XIII] activo de una forma din mica [AQ]

programme support actions [Programas de actuaci½n]#[13/95/51580100.P00] medidas de acompaÏamiento [JC]

project officer [VALUE]#[13/14490000/94] encargado de proyecto [JC]

prototyping [VALUE]#[13/14490000/94] construcci½n de prototipos [JC]

real life [VALUE]#[13/14490000/94. ... las dificultades que afrontan los procesos de innovaci½n en condiciones reales.] condiciones reales [JC]

Regional Innovation and Technology Transfer Strategies and Infrastructures (RITTS) [INNOVATION]#[13/95/53650000.P00] estrategias e infraestructuras regionales de innovaci½n y transferencia de tecnolog-a (RITTS) [ar]

Regional Technology Advisory Centres (RTAC) [SPRINT]#[13/95/53350100.P00] centros regionales de asesoramiento tecnol½gico (RTAC) [ar]

Regional Technology Plan (RTP) [INNOVATION]#[13/95/53650000.P00] Plan Tecnol½gico Regional (PTR) [ar]

relay centre [full form]#[Doc. es/13/94/16120100.p00] #centro de enlace [candemi]

Research and Technology Organisations (RTO) [full form]#[13/95/50660000.P00. Programa espec-fico para la difusi½n y optimizaci½n de los resultados de las actividades de investigaci½n y desarrollo tecnol½gico, incluida la demostraci½n (1994-1998)] organizaciones de investigaci½n y tecnolog-a (RTO) [candemi]

set-top box [full form]#[Doc. 13/95/5654 (adaptador multimedia interactivo: permite aprovechar la pantalla de un televisor dom_stico como monitor para la recepci½n de datos)] teleadaptador [candemi]

software [inform tica]#[13] programa (s)

sort code [banca]#[13/1257/94] c½digo de sucursal [AQ]

steering committee [radiocomunicaciones]#[13] comit_ de direcci½n [AQ]; comit_ directivo [AQ]

subsidy contract [full form]#[13/1257/94. Plan de trabajo] contrato de subvenci½n [JC]

success stories [full form]#[13] casos ejemplares [JC]

systems developer [full form]#[13] constructor de sistemas

target group [VALUE]#[13/14490000/94] grupo destinatario [JC]

technical specifications [licitaciones]#[13/95/51600200.p00. No confundir con las especificaciones t_cnicas, que son las caracter-sticas t_cnicas de un aparato (por ejemplo, velocidad, cilindrada, consumo, en un autom½vil)] condiciones t_cnicas [JC]; requisitos t_cnicos

technology performance financing (TPF) [SPRINT]#[13] financiaci½n de tecnolog-a en funci½n del rendimiento (TPF) [MC]

technology transfer project (TTP) [full form]#[13/50110000/95] proyecto de transferencia tecnol½gica (PTT) [mt]

technology validation project (TVP) [full form]#[13/50110000/95] proyecto de validaci½n tecnol½gica (PVT) [mt]

upstream [VALUE]#[13/14490000/94. Upstream integration...: integraci½n previa o ex ante de las perspectivas empresariales en los programas de IDT.] previo [JC]

F.2 Blind MARTIF to Eurodicautom conversion (source code on CD-ROM)

F.2.1 Output

```
%%BE ISOTC37
%%TY MTW96
%%NI 0
%%CF 1
%%EN
%%VE European Cooperative Research Action for Technology [1] CRAFT [2]
%%ES
%%VE Programa europeo de investigaci½n tecnol½gica en cooperativa [1] CRAFT [2]
%%BE ISOTC37
%%TY MTW96
%%NI 1
%%CF 1
%%EN
%%VE push-pull activity [1]
%%ES
%%VE actividad de empuje-tir½n [1]
%%BE ISOTC37
%%TY MTW96
%%NI 2
%%CF 1
%%EN
%%VE R&D Limited Partnerships [1]
%%ES
%%VE Sociedades Limitadas de I+D [1]
%%BE ISOTC37
%%TY MTW96
%%NI 3
%%CF 1
%%EN
%%VE task force [1]
%%ES
%%VE grupo operativo [1]
%%BE ISOTC37
%%TY MTW96
%%NI 4
%%CF 1
%%EN
%%VE awareness [1]
%%ES
%%VE concienciaci½n [1] conocimiento [2] sensibilizaci½n [3]
%%BE ISOTC37
%%TY MTW96
%%NI 5
%%CF 1
%%EN
%%VE back-up initiative [1]
%%ES
%%VE iniciativa de reserva [1]
%%BE ISOTC37
```

%%TY MTW96
%%NI 6
%%CF 1
%%EN
%%VE benchmarking [1]
%%ES
%%VE patrones de referencia [1] establecimiento de [2]
%%BE ISOTC37
%%TY MTW96
%%NI 7
%%CF 1
%%EN
%%VE bottom-up approach [1]
%%ES
%%VE enfoque ascendente [1]
%%BE ISOTC37
%%TY MTW96
%%NI 8
%%CF 1
%%EN
%%VE business plan [1]
%%ES
%%VE plan de negocio [1]
%%BE ISOTC37
%%TY MTW96
%%NI 9
%%CF 1
%%EN
%%VE Central Support Team [1] CST [2]
%%ES
%%VE Equipo central de apoyo [1] CST [2]
%%BE ISOTC37
%%TY MTW96
%%NI 10
%%CF 1
%%EN
%%VE Community Innovation Survey [1] CIS [2]
%%ES
%%VE Encuesta de la Innovaci½n Comunitaria [1]
%%BE ISOTC37
%%TY MTW96
%%NI 11
%%CF 1
%%EN
%%VE context driven application [1]
%%ES
%%VE aplicaci½n por contexto [1]
%%BE ISOTC37
%%TY MTW96
%%NI 12
%%CF 1
%%EN
%%VE demand pull [1]
%%ES
%%VE tir½n de la demanda [1]
%%BE ISOTC37

%%TY MTW96
%%NI 13
%%CF 1
%%EN
%%VE dissemination [1]
%%ES
%%VE difusi½n [1]
%%BE ISOTC37
%%TY MTW96
%%NI 14
%%CF 1
%%EN
%%VE distribution network [1]
%%ES
%%VE red de distribuci½n [1]
%%BE ISOTC37
%%TY MTW96
%%NI 15
%%CF 1
%%EN
%%VE European Association for the Transfer of Technology, Innovation and Industrial Information [1] TII [2]
%%ES
%%VE Asociaci½n Europea de Transferencia de Tecnolog-a, Innovaci½n e Informaci½n Industrial [1] TII [2]
%%BE ISOTC37
%%TY MTW96
%%NI 16
%%CF 1
%%EN
%%VE European Commission Host Organization [1] ECHO [2]
%%ES
%%VE Organizaci½n de distribuci½n de la Comisi½n Europea [1] ECHO [2]
%%BE ISOTC37
%%TY MTW96
%%NI 17
%%CF 1
%%EN
%%VE European Innovation Monitoring System [1] EIMS [2]
%%ES
%%VE Observatorio Europeo de la Innovaci½n [1] EIMS [2]
%%BE ISOTC37

F.3 Blind MARTIF to MultiTerm conversion (source code on CD-ROM)

F.3.1 Output

```
 **
<Created By>JC
<Personal Comment>IDT
<EN>European Cooperative Research Action for Technology
<TermTyp>Term
<EN>CRAFT
<TermTyp>Abrev.
<ES>Programa europeo de investigaci½n tecnol½gica en cooperativa
<TermTyp>Term
<ES>CRAFT
<TermTyp>Abrev.
**
<Created By>JC
<Personal Comment>VALUE
<EN>push-pull activity
<TermTyp>Term
<ES>actividad de empuje-tir½n
<TermTyp>Term
**
<Created By>ar
<Personal Comment>INNOVATION
<EN>R&D Limited Partnerships
<TermTyp>Term
<ES>Sociedades Limitadas de I+D
<TermTyp>Term
**
<Created By>JC
<Personal Comment>VALUE
<EN>task force
<TermTyp>Term
<ES>grupo operativo
<TermTyp>Term
**
<Created By>AQ
<Personal Comment>IMPACT
<EN>awareness
<TermTyp>Term
<ES>concienciaci½n
<TermTyp>Term
<ES>conocimiento
<TermTyp>Term
<ES>sensibilizaci½n
<TermTyp>Term
**
<Created By>JC
<Personal Comment>VALUE
<EN>back-up initiative
<TermTyp>Term
<ES>iniciativa de reserva
<TermTyp>Term
**
```

<Created By>candemi
<Personal Comment>13/95/50660000.P00
<EN>benchmarking
<TermTyp>Term
<ES>patrones de referencia
<TermTyp>Term
<ES>establecimiento de
<TermTyp>Abrev.
**
<Created By>JC
<Personal Comment>VALUE
<EN>bottom-up approach
<TermTyp>Term
<ES>enfoque ascendente
<TermTyp>Term
**
<Created By>JC
<Personal Comment>VALUE
<EN>business plan
<TermTyp>Term
<ES>plan de negocio
<TermTyp>Term
**
<Created By>JC
<Personal Comment>IMPACT
<EN>Central Support Team
<TermTyp>Term
<EN>CST
<TermTyp>Abrev.
<ES>Equipo central de apoyo
<TermTyp>Term
<ES>CST
<TermTyp>Abrev.
**
<Created By>aq
<Personal Comment>13/
<EN>Community Innovation Survey
<TermTyp>Term
<EN>CIS
<TermTyp>Abrev.
<ES>Encuesta de la Innovaci½n Comunitaria
<TermTyp>Term
**
<Created By>JC
<Personal Comment>VALUE
<EN>context driven application
<TermTyp>Term
<ES>aplicaci½n por contexto
<TermTyp>Term
**
<Created By>JC
<Personal Comment>VALUE
<EN>demand pull
<TermTyp>Term
<ES>tir½n de la demanda
<TermTyp>Term

**
<Created By>JC
<Personal Comment>VALUE
<EN>dissemination
<TermTyp>Term
<ES>difusi½n
<TermTyp>Term
**
<Created By>JC
<Personal Comment>VALUE
<EN>distribution network
<TermTyp>Term
<ES>red de distribuci½n
<TermTyp>Term
**
<Created By>JC
<Personal Comment>SPRINT
<EN>European Association for the Transfer of Technology, Innovation and Industrial Information
<TermTyp>Term
<EN>TII
<TermTyp>Abrev.
<ES>Asociaci½n Europea de Transferencia de Tecnolog-a, Innovaci½n e Informaci½n Industrial
<TermTyp>Term
<ES>TII
<TermTyp>Abrev.
**
<Created By>JC
<Personal Comment>IMPACT
<EN>European Commission Host Organization
<TermTyp>Term
<EN>ECHO
<TermTyp>Abrev.
<ES>Organizaci½n de distribuci½n de la Comisi½n Europea
<TermTyp>Term
<ES>ECHO
<TermTyp>Abrev.
**
<Created By>mt
<Personal Comment>13/94/16440400.P00
<EN>European Innovation Monitoring System
<TermTyp>Term
<EN>EIMS
<TermTyp>Abrev.
<ES>Observatorio Europeo de la Innovaci½n
<TermTyp>Term
<ES>EIMS
<TermTyp>Abrev.
**
<Created By>ar
<Personal Comment>INNOVATION
<EN>European Securities Dealers Association
<TermTyp>Term
<ES>Asociaci½n Europea de Intermediarios Financieros
<TermTyp>Term
**
<Created By>JC

F.4 Blind MARTIF to NACE conversion (source code on CD-ROM)

F.4.1 Output

```
<entry>
<EN>European Cooperative Research Action for Technology</>
<ES>Programa europeo de investigaci½n tecnol½gica en cooperativa</>
<CODE>0</>
</entry>

<entry>
<EN>push-pull activity</>
<ES>actividad de empuje-tir½n</>
<CODE>1</>
</entry>

<entry>
<EN>R&D Limited Partnerships</>
<ES>Sociedades Limitadas de I+D</>
<CODE>2</>
</entry>

<entry>
<EN>task force</>
<ES>grupo operativo</>
<CODE>3</>
</entry>

<entry>
<EN>awareness</>
<ES>concienciaci½n</>
<CODE>4</>
</entry>

<entry>
<EN>back-up initiative</>
<ES>iniciativa de reserva</>
<CODE>5</>
</entry>

<entry>
<EN>benchmarking</>
<ES>patrones de referencia</>
<CODE>6</>
</entry>

<entry>
<EN>bottom-up approach</>
<ES>enfoque ascendente</>
<CODE>7</>
</entry>

<entry>
<EN>business plan</>
<ES>plan de negocio</>
```

```
<CODE>8</>
</entry>

<entry>
<EN>Central Support Team</>
<ES>Equipo central de apoyo</>
<CODE>9</>
</entry>

<entry>
<EN>Community Innovation Survey</>
<ES>Encuesta de la Innovaci½n Comunitaria</>
<CODE>10</>
</entry>

<entry>
<EN>context driven application</>
<ES>aplicaci½n por contexto</>
<CODE>11</>
</entry>

<entry>
<EN>demand pull</>
<ES>tir½n de la demanda</>
<CODE>12</>
</entry>

<entry>
<EN>dissemination</>
<ES>difusi½n</>
<CODE>13</>
</entry>

<entry>
<EN>distribution network</>
<ES>red de distribuci½n</>
<CODE>14</>
</entry>

<entry>
<EN>European Association for the Transfer of Technology, Innovation and Industrial Information</>
<ES>Asociaci½n Europea de Transferencia de Tecnolog-a, Innovaci½n e Informaci½n Industrial</>
<CODE>15</>
</entry>

<entry>
<EN>European Commission Host Organization</>
<ES>Organizaci½n de distribuci½n de la Comisi½n Europea</>
<CODE>16</>
</entry>

<entry>
<EN>European Innovation Monitoring System</>
<ES>Observatorio Europeo de la Innovaci½n</>
<CODE>17</>
</entry>
```

```
<entry>
<EN>European Securities Dealers Association</>
<ES>Asociaci½n Europea de Intermediarios Financieros</>
<CODE>18</>
</entry>

<entry>
<EN>event</>
<ES>acto</>
<CODE>19</>
</entry>

<entry>
<EN>event schedule</>
<ES>programa de actividades</>
<CODE>20</>
</entry>

<entry>
<EN>ex ante</>
<ES>ex ante</>
<CODE>21</>
</entry>

<entry>
<EN>exploitation</>
<ES>explotaci½n</>
<CODE>22</>
</entry>
```

Appendix G: A Brief Argument for Careful Terminology Management

By Arle Lommel

Galinski, Christian. 1992. *International and regional co-operation in terminology*. <u>TAMA '92 Proceedings/Actes</u>. 3-20. Vienna: TermNet.

'Since the applications of terminologies have become so manifold lately, there is an increased necessity for co-ordination as well as for co-operative schemes in terminology.' (p. 4)

'It has been recognized that large amounts of terminological data cannot be managed by linguistic-lexicographical methods alone.' (p. 5)

'A number of UN organizations, such as the World Health Organization (WHO), the International Atomic Energy Agency (IAEA), the International Labour Office (ILO), the United Nations Educational, Scientific and Cultural
Organization (UNESCO), and the International Monetary Fund (IMF) are active in terminology, sometimes organizing networks of experts or institutions for terminological tasks.' (p. 7)

'Traditional means of access to information systems become more and more insufficient. Documentation languages, such as classification systems, thesauri, and indexing languages designed for access to knowledge contained in documents have to be supplemented by more sophisticated tools.' (p. 10)

'Ignoring the importance of terminology work leads to barriers in specialized communication since terminologies of low quality (containing many homonyms and synonyms) emerge and thus documentation tools remain inefficient.' (p. 11)

'Standardized terminologies (i.e., standardized systems of concepts to which standardized terms or graphical symbols are assigned) plan a key role both in scientific and technical communication as well as in information systems, documentation and knowledge engineering. Such terminologies can also be used as an essential tool for information management as well as for the transfer of information, knowledge, and technology.' (p. 12)

****

Galinski, Christian and Heribert Picht. 1996. *Graphic and other semiotic forms of knowledge representation in terminology management*. <u>TKE '96: Terminology and knowledge engineering</u>. 359-72. Vienna: INDEKS Verlag.

'Specialized communication always involves knowledge transfer, which in turn requires knowledge representation'

Generally the article argues for the inclusion of graphical data in term banks.

****

Schmitz, Klaus-Dirk. 1992. *Guidelines for the design and implementation of terminological data banks*. <u>TAMA '92 Proceedings/Actes</u>. 317-37. Vienna: TermNet.

'For scientific and technical editors and generally for all specialists involved [in] technical writing[,] such problems as synonymy, phraseology and other phenomena of special subject language as well as pragmatic aspects play an important role. The availability and application of terminological information is crucial both for the quality of a specialized text and for the consistency of its language and of its content.' (p. 324)

****

Ahmad, Khurshid, et al. 1996. *Engineering terminology-A case for a linguistically-informed terminology database*. <u>TKE '96: Terminology and knowledge engineering</u>. 166-78. Vienna: INDEKS Verlag.

'It can, perhaps, be argued that many a terminology database is created for the exclusive use of human translators or technical writers, and that the translator should, in principle, know the syntactic and morphological behavior of almost all the terms in the terminology database. Such an assumption regarding syntactic and morphological data is not really tenable where a terminology database is to be used as the primary lexical resource for, say, a machine translation or that of a grammar or spell checker uses systematically organized morphological data.

'Conventional terminology databases imply a 'flat' model of the lexical, syntactic, morphological, semantic and pragmatic data; the model is flat in that the data is stored at the most primitive level-merely as a string of characters-in tables of a relational database or as a network in a network database. Thus the interdependency of the various fields of data for a given term can only be exploited by a human user and not by a computer system'

****

Holmes-Higgin, Paul, and Khurshid Ahmad. 1996. *Is your terminology in safe hands? Data analysis, data modeling and term banks*. <u>TKE '96: Terminology and knowledge engineering</u>. 215-24. Vienna: INDEKS Verlag.

'The problem inherent in [storing terminology as flat files] is that [the terminology] is treated by the computer systems as no more than bits of text, meaningful only to the humans who placed them there; however, the current and future needs of terminology users asks (sic.) the computer to do more for them. Quality terminology is a key issue today. When the size of terminology collections becomes very large, the problems of insuring that data is not repeated in subtly different forms in different places are great.' (p. 215-16)

****

Eckman, Carol B. 1995. *The use of translation-oriented terminology methodology in the translation process*. <u>Translating and the computer 17</u>. (no consecutive numbering of pages). London: Aslib.

'In technical documentation of legal treatises, the quality of the text depends entirely on the consistent and correct use of a specific vocabulary, or terminology, Thus, one of the best ways to ensure quality is the record, standardize and circulate the terminology to be used.

'As translation work increasingly involves specialized texts, translators find themselves more and more in need of ways to record and, not least, to retrieve text-related information such as specialized terminology, source institution addresses, commonly used phrases of etiquette, names of organizations, etc.' (p 1-2 [numbering in article itself])

****

Goetschalkx, J. 1979. Eurodicautom. *Translating and the computer*, ed. By Barbara M. Snell, 71-75. Amsterdam: North-Holland Publishing Company.

Mere lists of equivalencies do not provide enough information.

'Equivalency is often at the level of the phrase and not necessarily between corresponding words.' (p. 72)

The article contains a list of administrative data categories in Eurodicautom that are necessary in the system.

****

Tanke, Eberhard H. 1979. *Implementing machine aids to translation*. <u>Translating and the computer</u>, ed. by Barbara M. Snell, 45-69. Amsterdam: North-Holland Publishing Company.

'It is obvious that special-subject languages differ from general language not only in their vocabulary, but also in their morphology, syntagmatic structures, phraseology, and, to some extent, their idiom. These features must therefore be considered in designing, developing, and implementing machine aids to translation. They are as essential to programming as to the acquisition of special-language data.' (p. 68)

****

Krommer-Benz, Magdalena. 1989. *International bibliography of terminological literature*. Vienna: TermNet.

276 page index of terminological data.

References

Dobson, R., Flohr, U., Graham, T., & Mace, S. (1998, March). "Weaving a better web." *Byte*. http://www.byte.com/art/9803/sec5/art1.htm

Hardman, D. (1996). *A Practical Proposal for the Blind Interchange of Terminological Data*. Provo, UT: Brigham Young University.

Hardman, D. (1997, June 10). *The Reltefä Specification*. http://www.ttt.org/clsframe/reltef/

Holmes-Higgen, P. & Ahmad, K. (1996). *Is your terminology in safe hands? Data analysis, data modeling and term banks*. Fourth International Congress on Terminology and Knowledge Engineering Proceedings. Frankfurt: Indeks Verlag.

Institute of Electrical and Electronics Engineers. (1990) *IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries.* New York, NY.

International Organization for Standards (ISO) (1998, October). http://www.iso.ch/

ISO 8879:1986, Information processing—text and office systems—standard generalized markup language (SGML). 1986. Geneva: ISO.

ISO 1087:1990, Terminology—Vocabulary. 1990. Geneva: ISO.

ISO FDIS 12620:1995, Terminology—computer applications—data categories. 1995. Geneva: ISO.

ISO FDIS 12200-1:1997, Terminology—computer applications—machine-readable terminology interchange format (MARTIF). 1997. Geneva: ISO.

Marchal, B. (1997, August 15). An Introduction to SGML: SGML in Plain English. http://www.pineapplesoft.com/reports/sgml/index.html

Melby, A. (1991). "MicroMATER: a proposed standard format for exchanging lexical/terminological data files." *META*, vol. 36:1, 135-160.

Melby, A. (1994a). "The translator workstation." *Professional issues for translators and interpreters* (vol. 7 in American Translators Association scholarly monograph series), ed. by Deanna L. Hammond, 127-147. Philadelphia: John Benjamins.

Melby, A. (1994b). "Terminology management in the SGML age." *Applications and implications of current LSP research: proceedings of the 9th european symposium on LSP, Bergen, Aug. 2-6, 1993.*, vol. 1., ed. by Magnar Brekke, Øivin Andersen, Trine Dahl and Johan Myking, 101-111. Bergen, Norway: Norwegian School of Economics and Business Administration.

Melby, A. (1995a). "Implementing the terminology interchange format." *Standardizing and harmonizing terminology: theory and practice, ASTM STP 1223*, ed. by Sue Ellen Wright and Richard A. Strehlow, 187-199. Philadelphia: ASTM.

Melby, A. (1995b). "E-TIF: an electronic terminology interchange format." *Computers and the humanities* (special issue: "The Text Encoding Initiative: background and context"), vol. 29:2, 159-165.

Melby, A. & Wright, S. E. (1998, October 19). *The CLS Framework*. http://www.ttt.org/clsframe/

Microsoft Corporation. (1998a, October 19). *Extensible Markup Language (XML)*. http://www.microsoft.com/workshop/c-frame.htm#/xml/default.asp

Microsoft Corporation. (1998b, December 16). *XML-Data*. http://www.microsoft.com/standards/xml/default.asp

Unicode Consortium. (1996) The Unicode Standard, Version 2.0. Reading, Massachusetts: Addison-Wesley.

Whatis.com.Inc. (1998, October 19). http://whatis.com/

World Wide Web Consortium. (1998a, October 19). *Extensible Markup Language (XML)*. http://www.w3.org/XML/

World Wide Web Consortium. (1998b, December 14). *XML Schema Working Group*. http://www.w3.org/XML/Activity.html

Wright, S. E. (1993). "Special problems in the exchange of terminological data." *TKE '93, Terminology and knowledge engineering*, ed. by Klaus-Dirk Schmitz, 300-307. Frankfurt: INDEKS-Verlag.

Wright, S. E. (1998). "Data Analysis for Data Modeling and Mapping Data Categories based on ISO 12200 and ISO 12620." *Proceedings of the 4th TermNet Symposium Terminology in Advanced Microcomputer Applications, "Tools for Multilingual Communication (TAMA '98)*. Vienna: TermNet.

Glossary

blind interchange:	Details of the interchange format are pre-defined so that export and import routines can be written without "seeing" the interchange partner.
CLS Framework:	Also known as *Concept-oriented with Links and Shared references Framework*, the CLS Framework deals with the structure of terminological databases and with information interchange between terminological databases.
conceptEntry:	A terminological entry identified by a concept identifier that defines a specific concept and lists the terms associated with that concept.
data model:	By data model it is not meant a type of data model such as entity-relationship diagrams in general, but a particular conceptual data model within some approach.
datcat:	Short for *data category*, a datcat is an element within a meta-language that describes the content of a particular tag.
DCS file	Short for *data category set*, the refObjectList structure for the data category.
document type definition:	A document type definition (DTD) is a specific definition that follows the rules of the Standard Generalized Markup Language (SGML). A DTD is a specification that accompanies a document and identifies what the markup is that, for example, separates paragraphs and identifies topic headings. By including a DTD with a document, any location that has a DTD "reader" (or "SGML compiler") will be able to process the document and display or print it as intended. This means that a single standard SGML compiler can serve many different kinds of documents that use a range of different markup codes and related meanings. The compiler looks at the DTD and then prints or displays the document accordingly. (Whatis, 1998)
element attribute:	A data category specification for each data category, indicating the data type associated with that item.
encodingDesc:	Short for *encoding description*, the encodingDesc is part of the header information of a Blind MARTIF document and contains information about how the terminological data is marked up, including what DTD is used.
Eurodicatom	*ECO* for short throughout this thesis, Eurodicautom is the multilingual terminological database of the European Commission's Translation Service.
Final Draft International Standard:	Also known as FDIS, a Final Draft International Standard is a draft of a work that is proposed to technical committees within ISO for approval as an international standard. Publication as an international standard requires approval by at least 75% of the member bodies casting a vote (ISO FDIS 12620:1995).
GlobeDisk	*GBD* for short throughout this thesis, GlobeDisk is a set of programs that provide spellchecking, thesaurus, dictionary, and translation capabilities, including terminology databases for translation between

	many European language pairs.
homograph:	A word that is spelled like another, but that has a different pronunciation, meaning, and origin. (ISO 12620:1995)
HTML:	HTML (Hypertext Markup Language) is the set of markup symbols or codes inserted in a file intended for display on a World Wide Web browser. The markup tells the Web browser how to display a Web page's words and images for the user. (Whatis, 1998)
International Organization for Standards:	Also known as ISO, the International Organization for Standardization is a worldwide federation of national standards bodies from about 130 countries, one from each country. The mission of ISO is to promote the development of standardization and related activities in the world with a view to facilitating the international exchange of goods and services, and to developing cooperation in the spheres of intellectual, scientific, technological and economic activity. (ISO, 1998)
interoperability:	The ability of two or more systems or components to exchange information and to use the information that has been exchanged.
langSet:	Short for *language set*, a langSet is a symbol used to designate the name of a language associated with a *conceptEntry*. The symbols used are from ISO 639 and are usually 2-letter symbols.
loss of information:	Related to *preservation of information*, loss of information refers to information not passed through in an interchange operation from one format to another.
MARTIF	MARTIF, short for *MAchine-Readable Terminology Interchange Format*, is a format to facilitate the interchange of terminological data among terminology management systems. This format is the result of several years of intense international collaboration among terminologists and database experts from various organizations, including academic institutions, the Text Encoding Initiative (TEI), and the Localisation Industry Standards Association (LISA). MARTIF, also know as ISO 12200, is associated with ISO Technical Committee 37 (Melby, 1998).
maxi-set:	The maxi-set of data categories is the total set of data categories (from ISO FDIS 12620) allowed in Blind MARTIF.
meta-language:	A language that describes another language. Meta means "having a higher stage of development." SGML and HTML are meta-languages because one uses them to describe other languages, i.e., they are mark-up languages. (Whatis, 1998)
MultiTerm	MultiTerm is a flexible and customizable database format for storing and managing terminological data. The fortmat accommodates many types of requirements, including flexibility in the usage of data categories within the database.
NACE	A simple terminological database, specifically a 7-language nomenclature file in ANSI character coding with SGML-style tags.
negotiated interchange:	Two or more partners use a common framework for interchange and negotiate details of the intermediate format to allow the writing of export and import routines that preserve as much information as

	possible.
ntig:	Short for *nested term information group*, a *ntig* is a grouping of data categories containing information about a termEntry, including *termGrp* and *auxInfo* (see Appendix A for structure).
parse:	In SGML-encoded documents, the process of comparing actual structure and markup to declared standards for the document in order to determine whether the document is valid.
parser:	A parser is a program that receives input in the form of sequential source program instructions, interactive online commands, markup tags, or some other defined interface and breaks them up into parts (for example, the nouns (objects), verbs (methods), and their attributes or options) that can then be managed by other programming (for example, other components in a compiler). A parser may also check to see that all input has been provided that is necessary. (Whatis, 1998)
preservation of information:	Whereas *loss of information* is to be avoided in terminology interchange, preservation of information is the goal, i.e. striving to pass on as much information as possible in a conversion from one format to another.
ptr:	At type of data category that points to other, related information within a MARTIF document.
refID:	A unique identifier (resource identifier) of a refObject (see refObject) found in *sharedRef*s (see *sharedRef*s).
refObject:	Found in the *sharedRef*s, a refObject is a unique set of identifiers and characteristics pertaining to the terminological data in the body of a MARTIF document.
Reltef™:	Reltef™ is an implementation of the CLS Framework. It includes an entity relationship diagram and can be implemented in any medium range to high-end relational database management system.
Rich Text Format (RTF):	RTF (Rich Text Format) is a file format that lets one exchange text files between different word processors in different operating systems. In some cases, the RTF capability may be built into the word processor. In others, a separate reader or writer may be required. (Whatis, 1998)
SGML:	Also know as the Standard Generalized Markup Language (SGML) ISO 8879:1986. A document that is represented as a sequence of characters, organized physically into an entity structure and logically into an element structure, as described in ISO 8879. An SGML document consists of data characters, which represent its information content, and markup characters, which represent the structure of this data and other information useful for processing it. In particular, the markup describes at least one document type definition, and an instance of a structure conforming to the definition. (Whatis, 1998)
sharedRefs:	Associated with a terminology database, *sharedRef*s are a set of references within a MARTIF document that can be shared among multiple *ConceptEntry*s or parts of *ConceptEntry*s. For example, a number of *ConceptEntry*s may be created by the same terminologist.

	That information can be represented in the *sharedRef*s, with each of the *ConceptEntry*s pointing to a common refObject.
source:	In a conversion from one format to another (e.g. one language to another) the source is the original format. For example, in a translation from German to English, German is the source language.
target:	In a conversion from one format to another (e.g. one language to another) the target is the format into which the conversion takes place. For example, in a translation from German to English, English is the target language.
termbase:	Short for *terminological database*, a termbase is a collection of terms and a possible description about the relationships between the respective terms.
terminology management:	The collection, organization and distribution of terminology. The goal of terminology management may be to encourage consistency within translation.
tig	Short for *term information group*, the value of this data category has diminished in favor of using the *ntig* (see ntig definition above).
Unicode:	Unicode is a system for "the interchange, processing, and display of the written texts of the diverse languages of the modern world." It also supports many classical and historical texts in a number of languages. Currently, the Unicode standard contains 34,168 distinct coded characters derived from 24 supported language scripts. These characters cover the principal written languages of the world. (Whatis, 1998)
XML:	XML (Extensible Markup Language) is a flexible way to create information formats and share both the format and the data on the World Wide Web, intranets, and elsewhere. For example, computer makers might agree on a standard or common way to describe the information about a computer product (processor speed, memory size, and so forth) and then describe the product information format with XML. Such a standard way of describing data would enable a user to send an intelligent agent (a program) to each computer maker's Web site, gather data, and then make a valid comparison. XML can be used by any individual or group of individuals or companies that wants to share information in a consistent way. (Whatis, 1998)