ISO Organisation internationale de normalisation

ISO TC 37/SC 4 N088

2003-10-30

Language Resource Management

Descriptors and Mechanisms for Language Resources

File ID SC4N088 doc (161 ko)

SC4N088.pdf (214 ko)

Title: Terminology and other language resources — Lexical Resource Markup Framework (LMF)

Editor(s): Monte George

Source: WG 1

Project number: N/A To be attributed after WI registration by ISO CS - This reference will supersede all previous one and remain attached as working ref along the duration of the project.

Status: Draft document to be attached to NWI letter ballot

Date: 2003-10-30

Agenda / Action: For transmission to ISO SC

References:

Contents

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

International Standard QQQQQ was prepared by Technical Committee ISO/TC 37, *Terminology (principles and coordination)*, Subcommittee SC 4, *Language resource management*.

ISO QQQQQ is designed to coordinate closely with ISO CD 12620-1, *Terminology and other language resources – Part 1: Specification of data categories and management of a data category registry for language resources*, *ISO CD 12620-3, Terminology and other language resources – Part 3: Data* categories for electronic lexical resources (ELR),and ISO DIS 16642, *Computer applications in terminology – TMF (Terminological Markup Framework)*.

Annex A is for information only.

Introduction

0.1 General aim of the standard

Lexical resources created in electronic form, in many cases outside the framework of lexicographical dictionaries designed for presentation in print form, abound throughout the language industry. Although some follow recognized procedures for the creation of lexicographical or even terminological works, many are generated in individual enterprise environments specifically to meet the needs of small work groups and thus demonstrate a high degree of variability. Most of these collections follow a generally lexicographical data model (i.e., each entry consists of one lexical unit or lexeme associated with potentially multiple senses and, in the case of multilingual resources, their respective bilingual or multi-lingual equivalents), yet there is no standard format for structures and the choice of the data categories treated from resource to resource varies considerably. The knowledge residing in such resources is extensive, and the desire to merge disparate lexical resources into larger global repositories is very strong, but standards are required in order to facilitate the creation of such resources via merging and interoperability.

0.2 Relevant International Standards

For terminology activities and management in general, the following International Standards are relevant: ISO 704, ISO 860, ISO 1087, ISO 10241.

1 Scope

This standard describes a high level model for representing data in lexical resources used with multilingual computer applications. The goals are to support the development of large scale multilingual lexical databases, to manage the exchange of data between lexical and terminological resources, and to provide a common model for creating lexical data that can be shared across different lexical and terminological resources, including machine translation lexicons. This standard is designed to be used in close conjunction with the metamodel presented in ISO 16642: Terminology Markup Framework and with ISO 12620-3:200?, *Terminology and other language resources — Data categories for electronic lexical resources (ELR).*

2 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of ISO QQQQQ. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on ISO QQQQQ are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

ISO 639-1:2002, *Codes for the representation of names of languages – Part 1: Alpha-2 Code*.

ISO 639-2:1998, *Codes for the representation of languages – Part 2: Alpha-3 Code*.

ISO 639-3:200?, *Codes for the representation of languages – Part 3: Alpha-3 Code for the comprehensive coverage of languages.*

ISO 12620-1:2000?, *Terminology and other language resources – Part 1: Defining parameters for specifying data categories for terminology collections and other data resources.*

ISO 12620-2:200?, *Terminology and other language resources – Part 2: Data categories for electronic terminological resources (ETR).*

ISO 12620-3:200?, *Terminology and other language resources – Part 3: Data categories for electronic lexical resources (ELR).*

ISO 3166-1:1997, Code for the representation of names of countries and their subdivisions – Part 1: Country codes.

ISO 3166-2:1998, *Code for the representation of names of countries and their subdivisions – Part 2: Country subdivision code.*

ISO 3166-3:1999, *Code for the representation of names of countries and their subdivisions – Part 3: Codes for formerly used names of countries.*

ISO 8879:1986, *(SGML) as extended by TC2 (ISO/IEC JTC 1/SC 34 N 029:1998-12-06) to allow for XML*.

ISO/IEC 10646-1:2000, *Information technology – Universal multiple-octet coded character set (ucs) – Part 1: architecture and basic multilingual plane.*

ISO/IEC 11179-3:2003, *Information Technology – Data management and interchange – Metadata Registries (MDR) – Part 3: Registry Metamodel (MDR3)*

ISO 15924:200?, Code for the representation of names of scripts

ISO 16642:2003, *Computer applications in terminology – TMF (Terminological Markup Framework)*.

3 Definitions

For the purposes of this International Standard, the terms and definitions given in ISO 1087-1, ISO 1087-2 and the following apply:

electronic lexical resource
ELR
database collection consisting of individual data entries each of which documents a lexeme and provides data pertinent to the senses associated with that lexeme, as well as in some cases equivalent terms or lexemes in one or more languages

electronic terminological resource
ETR
database collection consisting of individual data entries each of which documents a concept and provides data pertinent to the terms associated with that concept in one or more languages

machine translation lexicon
lexical resource in which the individual entries contain equivalents in two or more languages together with semantic information to facilitate automatic or semi-automatic processing of lexical units during machine translation

4 The use of terminological and lexical resources

Technical communicators and other authors, translators, language students, documentation specialists, readers, and analysts make use of a wide variety of research tools to improve their understanding of the topics and materials they are trying to understand, to express, and to translate. Traditional research tools have included print dictionaries, terminological resources, concordances, grammatical reference works, morphological tables, previously written related documents, and, in the case of translators, previously translated text files. The range of the research materials needed for a specific task depends on a number of factors, including the individuals' skill levels, their degree of specialization, the nature of the material to be processed, and the objective of the task in question. International business communications; documentation for computer software, pharmaceuticals, and other Industrial products; as well as literary works and on-the-spot support for peacekeeping and humanitarian efforts all produce unique challenges for writers and translators and require different authoring approaches and modes of translation.

Over the past several decades, on-line lexical resources have focused on reducing the significant amount of time spent looking up specialized terminology in print resources and have led to improved accuracy and consistency in the use of both terms and other lexical units. The success of these efforts to reduce turn-around times and to improve the accuracy and consistency of language products has led to the development of a wide variety of both terminological and lexical databases.

- Terminological resources provide critical support to the localization industry, as well as in a wide range of technical environments, where access to specialized vocabulary is needed, where authoring and translation tasks are focused on particular domains, source materials are generally expository in nature, and authors and translators possess high skill levels.

- Lexical resources are needed to support authoring and translation tasks where there is a need to understand the wider linguistic context of the source materials, where authors and translators may have lower skill levels, and where the subject matter covers a wider variety of subject domains. Lexical resources are also needed to support machine-assisted translation systems, part-of-speech taggers, and natural language processing systems, all of which require data with a full range of linguistic features. Finally, lexical resources are useful to end-users of information resources simply for the purpose of understanding and analysis.

While both types of databases serve important roles in different sectors of the language industry, electronic terminological resources (LTRs) have become by far the most important resource serving authoring and translation work groups in the industrial, commercial, and public sectors. LTRs have gained prominence due in large part to the role they play as a primary resource for the localization industry, which is the largest sector in the translation industry in terms of market share. They are ideally suited for authoring and translating expository documents that are rich in specialized vocabulary, such as international marketing materials, technical documentation for large scale construction and engineering projects, and pharmaceutical documentation, where poorly written materials and mistranslations can lead to the death or serious injury of product consumers.

Nevertheless, lexical resources are a constant mainstay across many different applications because of their great variety in form and content. The term "lexical resource" (which includes electronic lexicons and dictionaries of most kinds) describes a wide variety of research materials, which can include lemma-based lexicons, as well as resources documenting foreign phrases, place names, and personal names, and not excluding termbases. Lexical resources can be characterized according to three categories: range, perspective, and presentation:

- *Range* refers to the size and scope of the lexical resource, including the number of languages covered (monolingual, bilingual, multilingual).
- *Perspective* covers a number of factors, including the organization of the lexical resource (by concept, alphabetically, by morphological root, etc.).
- *Presentation* includes the format, the complexity of the documentation, and the scope of features (pronunciation, etymology, examples of usage).

As a result of differences in these factors, print dictionaries and electronic lexical rexources (ELRs) are highly variable in structure and content. This variety becomes more pronounced when linguistic differences are factored in across languages. Another aspect that increases the complexity of on-line resources is the common use of run-on entries in print dictionaries, both as a means of saving valuable space, and in order to link related and derived forms of the word to the key used for the dictionary look-up or to the search units used in ELRs.

5 Standards development for lexical resources

One major factor in the success of computer-based terminological databases is their well-defined structure based on a conceptual rather that a lexical framework, as exemplified in ISO 16642:2003. Multilingual terminological resources comprise a subset of the lexical resource genre with respect to range, perspective, and presentation. As a subset, ELRs can be implemented in coordination with the growing use of taxonomies and enterprise-specific ontologies organized on a conceptual framework that is highly suited for information management in computerized environments. Increasingly, ELRs have been integrated into a variety of other communications and translation support systems, including machine translation systems, translation memory, and controlled authoring systems.

In contrast, the complexity of lexicographical resources and the conventions of the printed dictionary have limited the ability to adapt print dictionaries in particular to processable on-line databases. Most efforts to develop models for encoding dictionaries have reflected the complex hierarchical structures of print dictionaries. Despite a general recognition that this degree of complexity conflicts with the constraints required by the database view, a large number of dictionary models have sought to support both views. This may be partly due to the strong desire to use information technology as a means to better automate the preparation of dictionary presentations and layouts for the publishing industry. As a result, current dictionary models do not adequately support the development of ELRs and applications designed primarily for use as on-line reference tools and for dynamic automatic processing.

A number of standards for terminology interchange and interoperability have emerged, such as ISO 12200:1999 and 12620:1999, and industry standards such as the LISA TBX format, which is an XML standard mostly compliant with SGML-based ISO 12200:1999. These standards, however, are derived from the concept-based structure of standard terminological resources, which provides a degree of similarity that has supported harmonization under ISO 16642, the Terminology Markup Framework (TMF). The markup formats that have emerged to support lexicography are more divergent at the metamodel level than the terminological standards due to the greater diversity in dictionary formats and the lack of a unified definition of the ultimate purpose of the standards. As noted, the current dictionary standards are focused primarily on the automation of print dictionary presentation and layouts. This international standard addresses two issues:

1. The need for a lexical resource model that will facilitate the development of on-line lexical databases focused on providing support for a range of authoring, translation, and information management activities for all sectors of the language industry that need lexical resources;
2. A Lexical Markup Framework (LMF) to serve as a metamodel for the development of markup languages for lexical resources suited for on-line research tools .

In addition, existing lexicographical exchange formats do not address the exchange of data between ELRs designed primarily for data processing and ETRs, nor are they compatible with existing standards for xml-based interchange of terminological information (TBX and OLIF). Their structures are not well suited to the development of enterprise translation systems that incorporate bilingual and multilingual electronic resources in multiple languages and formats. Enterprise systems require more constrained data structures to support the storage, processing, and retrieval of lexical data. Standards that provide developers and publishers with tools for tailoring the format of their lexical data for specific presentational products permit the emergence of a wide range of allowable data formats. Converting data between such an extensive assortment of formats becomes a non-trivial process that requires labor intensive manual conversion processes.

LMF will provide a common structure for data content management and the exchange of data between processable ELRs and ETRs (see section 9). This common structure is achieved by constraining the LMF data structure on the core lexical structure, the lexeme, and reducing the associative links between lexemes to a single information object (the refForm). By separating data content management functions from data presentation functions, LMF also facilitates the exchange of lexical data by constraining processing routines to the manipulation of content features, which is in keeping with the core philosophy of XML and that of evolving information management environments such as the Semantic Web.

6 The benefits of LMF

Without a model for handling both content management and presentation management, the lack of consistency in the data formats used for lexical resources inhibits the development of commercial markets for these resources and prevents the development of effective, large scale computer applications for the language industry. This will also restrict the market for lexical databases that customers can integrate into larger enterprise systems, reducing markets for vendor products.

In contrast to lexical resources, the development of on-line terminological resources has not been inhibited by this constraint because of their well-defined and consistent data structure. LMF will provide a common structure for creating, managing, and marketing lexical data resources. By separating the data content management functions from the data presentation functions, LMF will allow vendors greater flexibility in developing and marketing their products, for instance:

- Data in LMF can be licensed directly to end users for integration into authoring and translation systems or be acquired by publishers from data developers, thus encouraging the dissemination of lexical information.
- Publishers can license either data or database components to end users for plug and play integration into enterprise-based authoring and translation systems, opening new markets.
- Publishers can use a common format for acquiring new lexical resources from suppliers, encouraging the growth of the supplier base.
- Publishers will have a common data format that can be readily used to export lexical and terminological data to a variety of print dictionary formats, thus reducing publication costs.

7 The LMF metamodel

Lexical resources are comprised of lexical entries. LMF describes a lexical entry as treating a lexeme in the broadest sense: a linguistic unit belonging to a specific syntactic category with a particular set of senses. At the highest level, the entry is accompanied by global information that pertains to the entire entry, much of which is likely to be administrative in nature.

The entry itself consists of a form-sense pair or pairs. The form-sense pair is represented in the model by the "lexeme” node, which functions as a container comprising the *form* (represented here by the *keyForm*) and the *sense*. Each of these components of the form-sense pair will be accompanied by its respective data categories in the lexical entry. For instance, the keyForm is frequently accompanied by data specifying the syntactic category (part of speech) and other principal attributes of the lexeme (e.g. etymology, language, etc.), whereas the attributes associated with the sense may include the subject field, definition, contextual information, etc., as well as zero to many related lexical forms (*refForms*, e.g., synonyms, derived forms, etc.). These relationships are illustrated in figure 1.

The keyForm, the representational element in the form-sense pair, shall contain any instance of a wide range of linguistic units typical of dictionaries (a word, phrase, personal name, place name, or possibly even a symbol or a formula) and is not limited to lemma-based lexicographical elements. Nevertheless, it is easier to understand the function of the keyForm if one bears in mind the fact that keyForms used in print dictionaries are often called "head words", although this designation is misleading (not all keyForms are just words) and ambiguous (the term "head word" is sometimes used to refer to the nucleus of a multiword lexical unit or compound). Note that there shall be one and only one instance of a keyForm in a given lexical entry. In contrast to concept-oriented terminological entries that treat a concept and its many representational forms, a lexical entry treats a single keyForm associated with the lexeme container, its potentially many senses, and all the descriptive and administrative information associated with that keyForm.

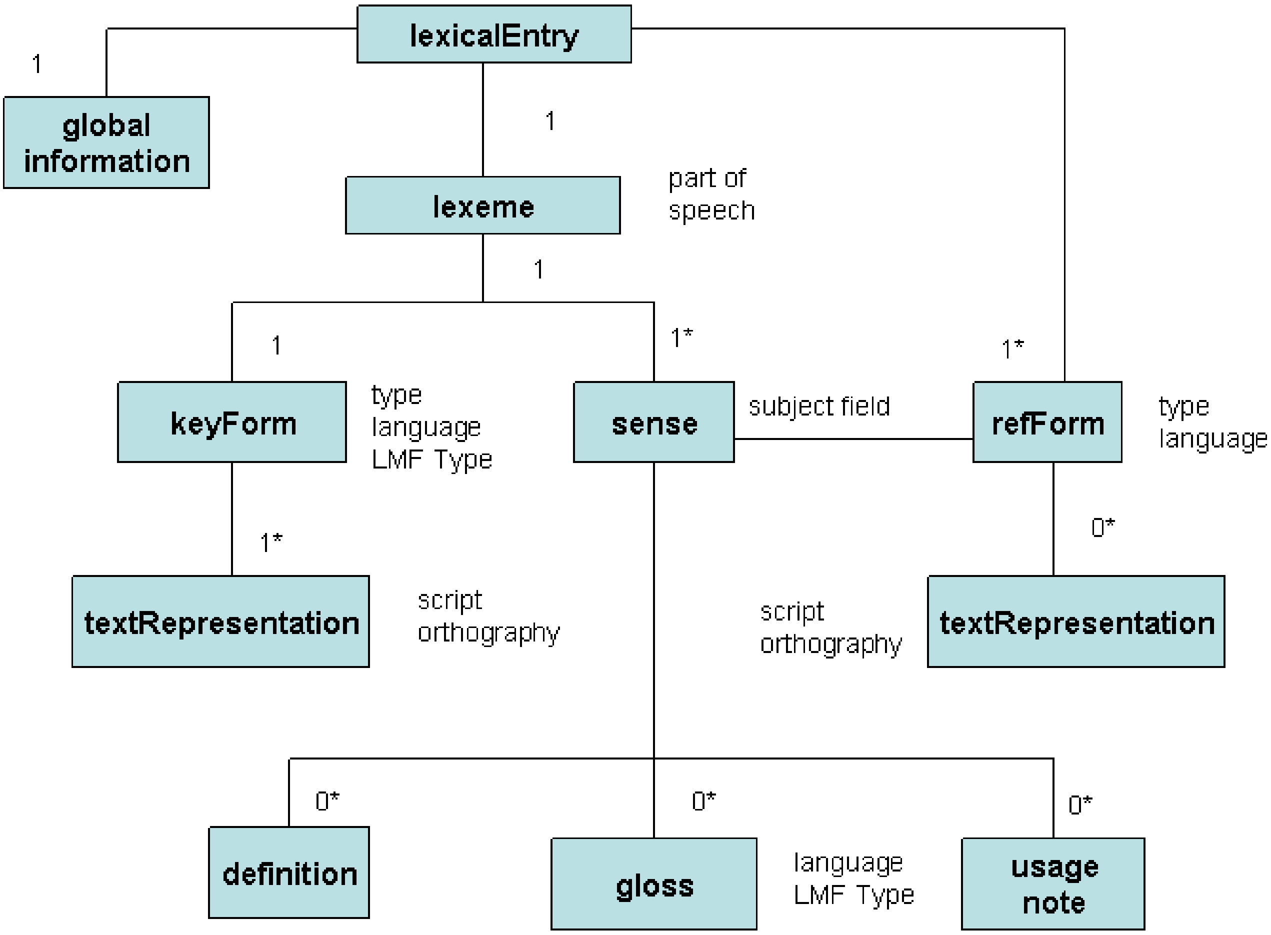

Figure 1: The LMF metamodel

The sense component contains concept-related information (a definition and subject field information) and additional data categories (bilingual and multilingual equivalents in other languages, usage notes, etc.). The information included in the sense component is identified by various data categories used to identify the different kinds of descriptive values associated with the keyForm.

Each of the zero to many refForm elements contains a single linguistic unit (word, phrase, personal name, place name, etc.) related in some aspect to the keyForm. The metadata reference (e.g., the data category) associated with any given refForm describes the aspect of the refForm relationship to the keyForm (synonym, root, derived form). The refForm can also explicitly point to an external keyForm whose semantic content is equivalent (either as a synonym or multilingual equivalent) to the conceptual content represented by the refForm. Here again, it is important to remember that although refForms *look like* keyForms, their role is very different in the lexical entry because they serve only to describe or provide additional metalinguistic information about the keyForm and one of the senses with which it is associated. Although many refForms (such as synonyms) are associated at the sense level, others (such as bases, Han characters, etc.) can be associated at the entry level. Hence the refForm is represented in the model as linked at either or both of these levels.

8 Metamodel components

8.1 Language related attributes

LMF provides a systematic method for identifying the language, script, and orthographies that distinguish the individual linguistic units in the *keyForm*, *refForm*, and *sense* elements of the lexical entry. This method of distinguishing the linguistic units enables the design of computer applications that are capable of handling bilingual or multilingual lexical resources in many languages. It also allows for the differentiation of different working and object languages used at virtually any level of the lexical entry. These kinds of multilingual data collections can contain many different source-target language pairs and even more complex associations of information objects in different languages. Specifying the language, script, and orthographical attributes of each linguistic unit enables the design of well-structured and efficient computer code for indexing, retrieving, and presenting data in multilingual processing systems. These methods are intended to enable the design of intuitive interfaces that relate to the user's view of information organization in multilingual dictionaries.

LMF uses three data categories to describe the representation of a linguistic unit (keyForms, refForms, etc.): 1) the language attribute (expressed as xml:lang), 2) a writing system attribute (including scripts and variant writing systems, such as pinyin for Chinese), and 3) an orthography attribute (variations in spelling, transcription, or transliteration). LMF uses ISO 639 codes for the language and ISO 15924 (where relevant) for scripts and writing systems. The orthography element generally contains user-defined codes due to the frequent occurrence of domain specific data and the current lack of standard codes for orthographies.

8.2 The keyForm

As noted, the keyForm is the representational form in the form-sense pair that comprises the core of the lexical entry and is one component of the content of the lexeme container. The keyForm comprises a designation, usually, but not always, in linguistic form, used for communication and thought in conjunction with the more abstract sense, which comprises the other compoent of the lexeme container. In terminology theory, the sense is equatable to the concept associated with a given lexeme in a given subject field or situational context. The keyForm contains concrete representations of the designation in different media. This international standard is limited to a description of text representations and does not focus on other forms, such as icons, pictures, concrete objects, ideographs, etc. The textRepresentation category contains the variant orthographies that can be used to represent the form. Depending on the specific authoring or translation task involved, the standard orthography or any of a number variant orthographies may be best suited for representing the keyForm. For example, stenographic standards may be used to transcribe court records or standard transcriptions may be produced for end users who may not be able to read native script. In the keyForm or the refForm (see below), these variant orthographies are contained as a logical unit based on the linguistic form. This structure also logically groups writing systems for efficient computer implementation and information retrieval.

8.3 The refForm

A lexical entry contains zero to many refForms. Each refForm component is related to the keyForm by a specific type of association (e.g., derivative, synonym, antonym). The refForm contains one to many textRepresentations that directly inform the reader about grammatically related word forms. The textRepresentations provide a key for readers to use in querying for related lexical entries of interest, or as a link to other information resources, such as morphological tables. The refForm can have an external *crossReference* attribute, which explicitly links the refForm to another associated LexicalEntry, where (in some instances) the refForm is itself treated as the keyForm for that entry. These links can serve the purpose of hypertext-like information retrieval, but they can also be used to reconstruct virtual terminological entries based on the information contained in the lexical entry (e.g., tracing and assembling information related to synonym, abbreviation, and other related links).

8.4 The sense component

The sense component contains the definition, equivalents, usage notes (examples of usage), and general notes. The definition, equivalent, and usage notes can be associated with the language attributes set. Definitions can be in the source language (the language of the keyForm), and/or one to many target languages. (In this context, it is important to note that bilingual and multilingual lexical entries, with their single keyForms in a given language, are inherently directional with respect to source and target language, whereas many terminological systems are designed to be direction-neutral.) Equivalents can be provided for one or many target languages (bilingual vs. multilingual ELRs). Usage notes can be in the source language (monolingual resources) or one to many target languages (bilingual and multilingual resources). LMF also supports source-target usage pairs which consist of a sample usage or contextual reference in the source language and a translation of the sample usage in the target language. A attribute indicating degree of equivalence can be used to distinguish translations that are conceptual equivalents of the keyForm. This feature enables the transfer of equivalent terms in different languages between

LMF and TMF environments.

8.5 Relationship of orthographical variants to ISO 12620

LMF provides a flexible descriptive structure that allows computer end users to specify the language, writing system, and orthography to use. Frequently, variant orthographies correspond to more than one ISO 12620 data category and there is often a choice of transcriptions and transliterations available to translators. For example, a pinyin and tone orthography for Romanized Chinese (FEI1 JI1) can correspond to a transcription together with a syllabification.

9 Computer applications for interaction between electronic lexical and terminological resources

9.1 Lexeme-based structures

The LMF structure is based on the lexeme and constrained with respect to the primary attributes of the lexeme. LMF does not incorporate the deep hierarchical structure that is typical of print dictionaries. Although refForms can provide a limited set of run-on entries, and are informative in this regard, the true function of the refForm is to provide associative links between the keyForm that is the subject of a given entry and related lexical entries. Thus the refForm serves a descriptive, potentially hypertextual function. This constraint is a key factor in enabling the development of efficient computer applications that can provide intuitive user interfaces for bilingual applications containing many language pairs. Computer applications provide methods for retrieving, organizing, and presenting data that make the use of deep hierarchical structures used in print dictionary data unnecessary. Innovative computerized presentation of different sorts of information and different levels of the model can provide a “virtual” structure for well-defined data that eliminates the need to explicitly replicate print dictionary structures in the markup language. XML and UNICODE (ISO 10646-1:2000) have played a large role in making this approach practical. The model does not include the complex and highly variant features that characterize print dictionaries, but provides a constrained model based on the fundamental aspects of the lexeme – which is the core component of lexical resources, including print dictionaries, and which forms a bridge to terminological resources. Presentation and layout rules are handled outside LMF, but enabled through the well-defined LMF data structure, including the methods for associating the keyForms in the lexical data collection.

9.2 Conversion between LMF and TMF

TMF and LMF differ significantly in structure and content, restricting the exchange of data between the two models, but in ways that can be defined. LMF consists of broadly based linguistic units (nouns, verbs, phrases, proper nouns) which are associated with specific senses (conceptual content). In contrast, TMF is a concept-based model and contains a subset of the kinds of linguistic units found in LMF. Therefore, data exchanges between the two models will not be fully bidirectional. It should be possible to export all information objects from a TMF-based application to an LMF-based application, but only a subset of the information in LMF can be mapped to TMF. In addition, in order to map the data from TMF, an LMF implementation must explicitly distinguish which linguistic units in an entry are linked to a specific sense or concept, and which linguistic elements in the sense element are conceptual equivalents of the linguistic unit represented by the keyForm. For instance, for a given sense of a keyForm, a computerized linking routine could assume that synonyms, variants, orthographic options, abbreviations, and the like are all possible representations of a single concept. Information concerning these elements could be assembled, in some cases drawing on additional information contained in cross-referenced lexical entries, and expressed as a virtual terminological entry expressed in TMF. This kind of automatic generation of conceptual entries will never be perfect, however. For example, even a practical multilingual equivalent in the sense component may not provide a rigorous conceptual equivalent to the linguistic unit in the KeyForm. However, this process can be used to structure mutually compatible terminological and lexical views of the same data that will be highly reliable.

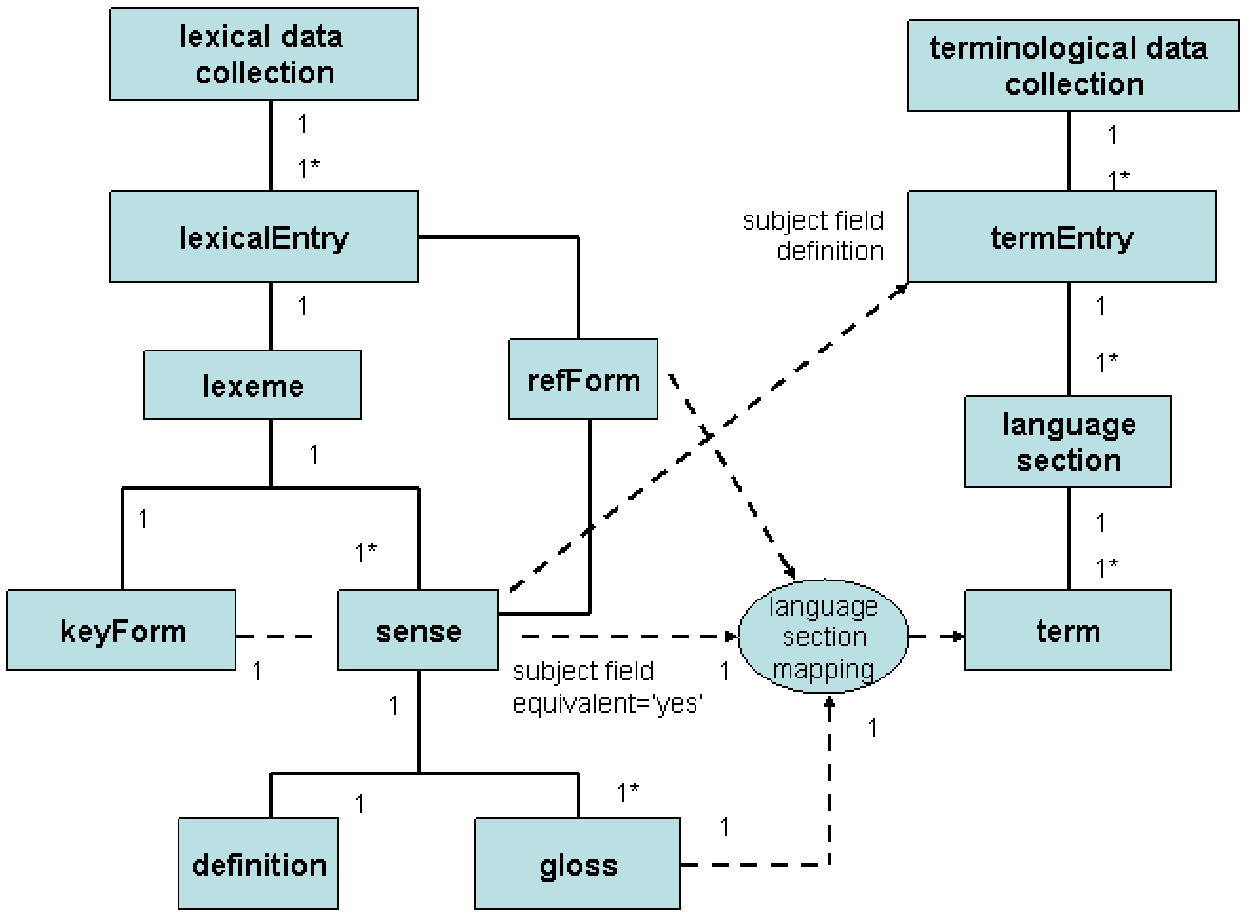

Figure 2: Exchange and interoperability between LMF and TMF

Figure 2 illustrates the relationships that shall exist in order to ensure exchange and interoperability between lexical and terminological resources. First of all, the subject field associated with a given sense in the lexical entry must match with or map to that of a given concept entry in the terminological entry. Secondly, the sense itself, as exemplified by its definition and other attributive data, must match to that of the concept treated in the terminological entry. Obviously, these matches can be difficult to guarantee if they are made automatically because some degree of polysemy can exist even in carefully constrained subject field collections. Given concept/sense equivalence, however, it can be assumed that the keyForm and any equivalents present in the lexical entry can be mapped to terms in the terminological entry, as can certain of the refForms (synonym, abbreviation, symbol, variant, orthography, etc.).

10 Conversion and interoperability between LMF and machine translation lexicons

[To be added.]

11 Conversion and interoperability between LMF and presentational formats for lexicographical resources

[To be added.]

12 Interoperability with ontologies and taxonomies (such as OWL)

[To be added.]

Annex A
Research Underlying this Document

This document is based on extensive analysis of the following data collections:

Languages	No. of Resources
Arabic	11
Azeri	1
Chinese	2
Hindi	1
Kirghiz	1
Kurdish	1
Macedonian	2
Pashto	1
Persian/Farsi	2
Punjabi	2
Russian	6
Croatian	2
Spanish	1
Tagalog	1
Turkmen	1
Urdu	3
Uzbek	2

Bibliography

[1] OLIF Open Lexicon Interchange Format http://www.olif.net/

[2] TBX. The TermBase eXchange format. http://www.lisa.org/tbx/

[3] UNICODE. The Unicode Standard 4.0. Boston: Addison-Wesley, 20003.

[4] W3C, Extensible Markup Language (XML) 1.0 (W3C recommendation 10-February-1998) (http://www.w3.org/TR/1998/REC-xml-19980210)

[5] W3C, XML Schema Part 1: Structures (W3C Working Draft 7 April 2000) (http://www.w3.org/TR/xmlschema-1)

[6] W3C, XML Schema Part 2: Datatypes (Working Draft 7 April 2000) (http://www.w3.org/TR/xmlschema-2)

[7] W3C, Resource Description Framework (RDF) Model and Syntax Specification: W3C Recommendation 22 February 1999. http://www.w3.org/TR/REC-rdf-syntax/

[8] Wright, Sue Ellen and Gerhard Budin. 1994. “Data Elements in Terminological Entries: An Empirical Approach”. Terminology. Vol. 1, No. 1. Amsterdam: John Benjamins Publishing Company.

[9] Wright, Sue Ellen. 2000. “Data Categories for Terminology Management”. In: The Handbook of Terminology Management. Vol. 2, Sue Ellen Wright and Gerhard Budin, compilers. Amsterdam: John Benjamins Publishing Company, 552-571.