Alan K. Melby, Brigham Young University, USA
Sue Ellen Wright, Kent State University, USA

LEVERAGING TERMINOLOGICAL DATA FOR USE IN CONJUNCTION WITH LEXICOGRAPHICAL RESOURCES

1 Integration Issues: Introduction

Integration is the operative term in today's localization (L10N) and translation environments. Translation departments and companies, as well as some individual translators, who have already grown accustomed to using workbench systems that combine translation memory (TM) with terminology management systems (TMS) are rapidly moving on to include an array of L10N utilities and machine translation (MT) in the mix. Furthermore, the evolution of electronic general language lexicographical dictionaries, artificial-intelligence-related ontologies, and other knowledge-based systems promises to add a supplemental dimension to the seamless mix of tools and resources available at the translator's workstation. Actually, the concept of integration is not new (Melby 1982), but widespread implementation is. In this move from island-like stand-alone applications (similar to the "islands of automation" seen a decade ago in computer integrated manufacturing), economic concerns and practical requirements have inspired system developers and users alike to seek solutions that will leverage information stored in existing concept-oriented terminological assets designed for use by human translators in order to employ this information in conjunction with lexicographical dictionaries, in particular MT lexica or other word-sense-oriented natural language processing (NLP) resources. Although MT lexica are not the only lexicographical databases involved, this article will generally reference them because market pressures have shifted the focus to these tools.

For purposes of this paper, the authors recognize that all the environments and traditions that we describe in the paper are working with terminology in one way or another, i.e., with technical or special languages. Thus it can be very confusing to try to contrast "terminological databases" with "lexicographical databases," as if somehow lexicographical resources did not also contain terms. Nevertheless, we need to have some means of distinguishing between the database tradition that has grown up around the General Theory of Terminology (Wüster 1973/1985) and the direction that the documentation of terms has taken in NLP and MT environments. In order to keep this all straight in a simple, shorthand sort of way, we will refer to the former databases as *termbases*, the latter as *lexbases*, and hybrid resources that leverage data from both kinds of database models as *lex/term-bases*.

With the completion and final approval of ISO 12200:1999, one of the next steps for ISO Technical Committee 37 on *Terminology (principles and coordination)* (ISO TC 37) is to amend the new standard to accommodate the importation and exportation of data between termbases and lexbases, as well as between and among various other language technology applications. This effort will take into account existing interchange formats such as those proposed by the LISA's OSCAR group (the Localisation Industry

Standards Association special interest group called “Open Standards for Container/Content Allowing Re-use”) and the OTELO group (Open Translation Environment for Localization) (OSCAR 1999; Carroll 1999; Thurmair et al. 1999).

Whatever tools mix is used, the synchronization of terminological resources is a critical procedure designed to ensure consistent terminology regardless of whether a given translation segment or unit was produced using TM or MT or written by a human translator using a termbase. Furthermore, no matter what the approach, terminology management is expensive and shared resources reduce labor-intensive duplication of effort.

Solutions involving the interaction of TM and termbase systems (e.g., between TRADOS MultiTerm™ and TRADOS Workbench™, or between TRANSIT™ and TermStar™) feature the integration of termbases with translation memory, but this type of coordinated lookup and data access does not require any change in the essential way the termbases run or the way data are represented. The programs are basically coordinated to function together in a user-friendly way, but human users can access the termbase data in more or less the normal way if necessary. They base their final translation decisions on information present in the TM corpus, on the textual material stored in the termbase (definitions, contexts, usage notes, transfer comments, etc.), or simply on intuitive human judgment. In contrast, the lexbases used for automated MT systems must rely on explicitly coded data that human translators already presumably know (e.g., morphological, syntactic, or semantic information) or that they can intuit based on the types of information just mentioned. Thus, although termbases lack important information needed by MT systems, they also possess a wealth of information that is inaccessible to automated systems. Hence, introducing termbase information either directly or as a dynamically shared resource presents problems that should not be downplayed, but that nevertheless should not be overestimated either.

2 Lexbase and Termbase Structures as a Function of Complexity

Creating multifunctional resources (our *lex/term-bases*) is not a new notion. Both Melby (1988) and Galinski (1988) predicted the utility of these kinds of assets and the ultimate cost savings to be derived from them. Sager (1990) provided a highly pragmatic, data-modeling orientation with guidelines on how to actually implement such databases, and Ahmad et al. (1996) described an actual functioning system (TRANSTERM) and made a plea for more widespread development of lex/term-bases. Nevertheless, some terminologists still question the feasibility of creating unified systems, while others unfamiliar with the differences between the data models in question are unaware of the challenges involved. Although many traditional terminologists take the concept-centric focus of their work for granted, the difference between the termbase and the lexbase approach should be reiterated. *In the beginning* may have been *the word*, but the beginning of terminology management within the termbase tradition lies, at least ostensibly, elsewhere. Wüster (1979/1985:1) unequivocally states that *every terminology activity is based on* ***concepts*** *and the distinction between them.* Consequently, the concept

and its multiple designators are the fundamental organizing feature for this type of terminology management (Felber 1984: 255).

Riggs described the difference between the lexicographical (traditional dictionary) approach and the terminological treatment of terms in concept-oriented records (evidenced in his INTERCOCTA glossary) as a function of the divergent treatment of polysemy and synonymy: (Riggs 1985:xiv):

- Lexicographical entries document words (or also technical terms) *and their (potentially) multiple polysemic senses.*
- Terminological entries document concepts *and their (potentially) synonymous designators (terms).*

This distinction between the traditional lexicographical (lexbase) entry and the terminological (termbase) entry is further explicitated in Wright-Budin (1997:328).

Both of these models for documenting lexical data reflect efforts to simplify the complexity inherent in language. In monolingual environments, the creation of standard dictionaries (a word with its several senses, as evidenced in multiple definitions, etc.) represents a manageable one-to-many relationship, but the addition of multiple languages into the mix introduces the potential for many-to-many relationships (many meanings coupled with many designators, which in turn may be associated with many other meanings in their own linguistic environments). One option for simplification is to eliminate lexicographical documentation (definitions, etc.) altogether and to merely record translation equivalents without descriptive information. We see this kind of approach in the glossary-style dictionaries that have proven to be both widespread and highly ineffective as translation tools.

The termbase approach provides a rich mix of data (definitions, contexts, transfer comments, etc), but it reduces potential complexity by confining itself to a single concept per entry and thus relegating polysemic units to separate terminological entries, i.e., polysemy doesn't disappear, except possibly in normative terminology, but it is dealt with by excluding it from the individual entry (Budin 1996). As noted above, MT environments do not necessarily benefit from the extensive body of information that can be supplied in a termbase entry because the text data accompanying a typical human-oriented term entry is inaccessible in a meaningful way to automatic parsing. Here it makes more sense to document one-to-one interlingual transfers or to treat headwords in a lexicographical way, differentiating among polysemic senses according to subject fields, clients, text types, etc. Data models for the lexbase and termbase approaches reflect these divergent solutions to the problem of complexity, and any effort to coordinate the various approaches must compensate for these fundamental differences.

3 Theoretical Structural Constraints Affecting Merging

When viewed strictly from a hard-copy representational perspective, the potential for integrating termbase with lexbase data may appear to be incompatible. Ordering

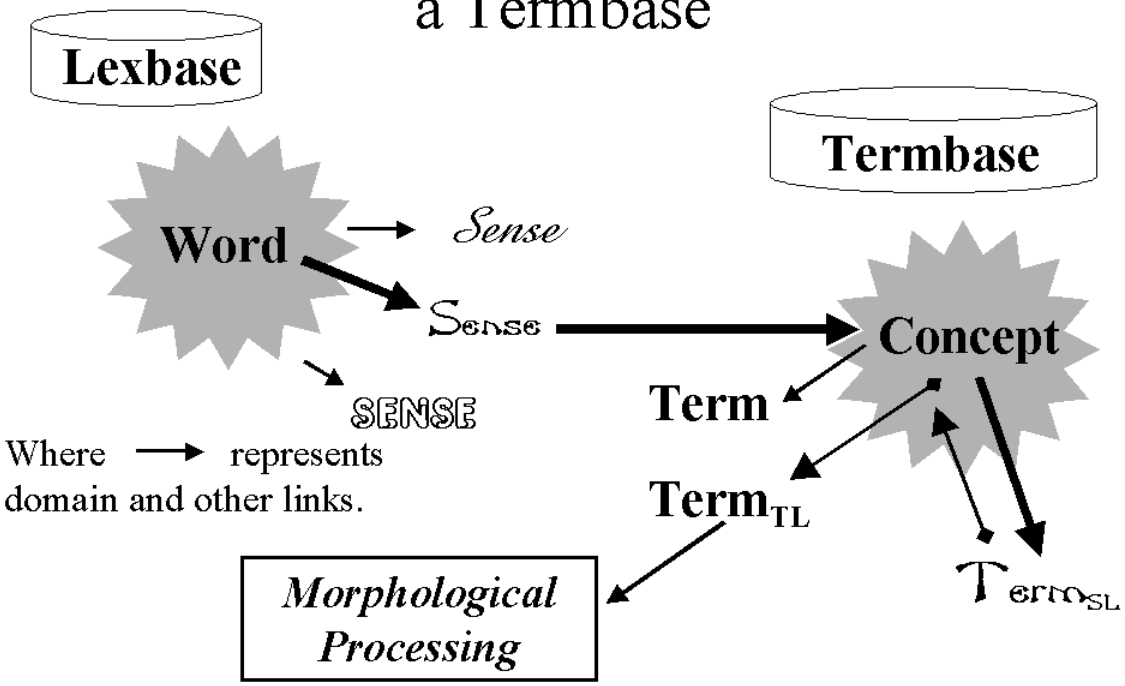
principles alone might prevent such merging: in print, lexbase resources feature standard alphabetical ordering, whereas many concept-oriented termbase collections are arranged according to concept hierarchies, which suggests the paper-oriented notion that it would be impossible to merge or otherwise coordinate the two kinds of systems. Of course, in modern computational environments, the order of words on the page is no longer relevant. As Sager has pointed out, in electronic collections "*terminological data can be collected regardless of the onomasiological* [word-oriented] *or the semiasiological* [concept-oriented] *approach since ordering of data occurs totally independently of compilation*" (Sager 1990:140). Modern termbase technology allows for multiple indexing and ordering, as well as the representation of multiple concept or classification systems within the same corpus of lexical data. Hence, the physical position of a concept or term within a given collection is probably only some sort of arbitrary address or identifier and no longer represents a durable feature of the meta-structure of a collection nor does it exert any kind of permanent effect on the representation of that data either in electronic or hardcopy form.

Figure 1 illustrates the structural constraints that characterize a dynamic interaction between an existing word-based lexbase entry and an existing concept-oriented termbase entry. The problems here are not necessarily the same as the problems involved in developing lex/term-bases "from scratch". In the best of all possible worlds, such as in the kind of controlled language environment that is ideal for implementing MT solutions, words and terms occurring in a text for translation might be univocal, i.e., monosemic and mononymic, thus eliminating complexity and rendering transparent the structural differences between lex and term entries. Indeed, the simpler the term entry or the lex entry, the fewer the apparent differences.

When working with typical texts and attempting to utilize existing lexical resources, however, polysemy and synonymy cannot generally be ruled out. Hence there can be multiple senses for any given word (the one-to-many relationship cited above), as exemplified by the various fonts on the lexbase side of the illustration. The presence of polysemy does not permit a single, automatic one-step link from a word in a lexbase entry to the term in a termbase entry that is needed to access the target language equivalent in that entry. Consequently, the figure postulates the presence of domain-specific or other kinds of markers (or explicit links) that facilitate the selection of the appropriate sense from among various senses of the word in question. Then, in order for an automatic search routine to navigate successfully from a source language word (SL) on the lexbase side of the equation to retrieve the appropriate target language equivalent (TL) on the termbase side, it would have to:

1 Recognize the word in the SL.
2 Identify the link for that word to one of potentially several specific domain or other usage-related references (optionally contained in the set-up instructions for a given MT job run, explicitly tagged in the text, or, in more sophisticated systems, self-generated based on such aspects as frequency or co-occurrence with other terms or collocates).

Figure 1: Navigating from a Lexbase to a Termbase



3 Identify the individual sense associated with that word/term in that domain or subdomain.
4 Link that specific sense to an entry for a domain-specific concept that has as one of its SL terms the same word or lemma represented by one form of the original word in the source text.
5 Via the implicit conceptual link afforded by the concept-oriented term entry, determine the appropriate target language term (represented by the angled arrows between $Term_{SL}$ and $Term_{TL}$ in Figure 1).
6 Prepare the term for insertion into the translated text using parsing rules and/or stored lexicographical data (which may or may not be present in the termbase), if necessary converting the retrieved lemma to the proper form for the TL term in context (shown here as a morphological processing unit, which may be a separate utility or be part of an MT system).

Needless to say, this procedure or sequence of procedures involves a complex threading mechanism. Certainly a word or term, or some form of it, is essential to initate the linking function: the word from the lexbase side of the process must match up at some level with a term on the termbase side. Since our illustration presupposes existing databases that were not created with the intent of interaction, there will be no hard-programmed link to achieve this connection. Of course, the linking element does not constitute a simple, one-piece key, not even when it is stripped of specific morphological features to arrive at some kind of stem-like lemma. If that were the case, complex terminology management

would be unnecessary and MT might probably have long since established itself as the translation method of choice. In the scenario described above, the lexbase entry is linked to the termbase entry by virtue of the following:

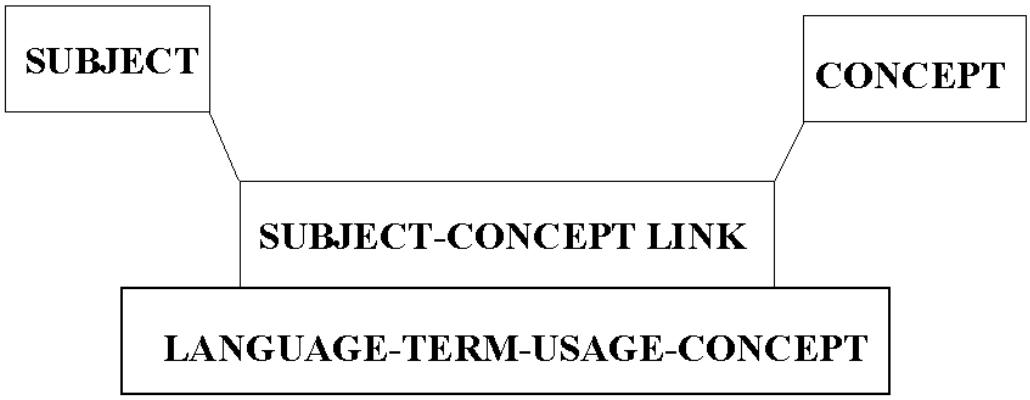
- On the lexbase side: a word + (probably) a part of speech notation + one or more domain references designed to clarify the sense of the word
- On the termbase side: a term that has the same form or an equivalent form at the lemma level as the lexbase word, that matches the part-of-speech (pos) notation, and that resides in a termbase entry devoted to the same domain or subdomain referenced in the lexbase (One difficult problem to solve at this level is the synchronization or harmonization of lemmatization processes in different applications.)

Even this three-part resolution (word/term + pos + domain reference) represents a blatant over-simplification. In an analysis of terminological equivalents in a single technical subdomain (threaded fasteners) for a single language pair (German-English), Anderson has observed a series of kinks in the chain that links a word in context with the appropriate concept-oriented target language term (Anderson 1999):

- Concept systems in the two languages are frequently asymmetrical.
- Standard parsing assumptions may have to be jettisoned to accommodate intertextual realities (e.g., the noun *bolt* may nicely become the verb *bolting*, but for *screw*, reflecting the serious audience expectations of a technical audience, one must avoid the obvious, but potentially ambiguous morphological solution and opt for the semantically safer *screwdriving* or even *tightening*). The same problem affects many possible collocates as well.
- Generic terminology for a domain (e.g., *threaded fasteners*) may be inappropriate when applied to a more specific subdomain (e.g., *automated fastening systems*): for example *Reinheitsgrad*, which can reference a *purity factor* when applied to fasteners themselves, becomes *cleanliness and dryness*, or recategorized, *clean and dry*, when used to characterize batches of fasteners intended for use in an automatic feed system.
- Terminology that is appropriate for one manufacturer may be totally unacceptable to a competitor in the same field.
- Client-side polysemy may make it necessary to further differentiate definitions and equivalents for different departments or divisions of the same company.
- Terminology and collocational usage that are acceptable in North America may not be appropriate or even comprehensible in Britain, Australia, or India.

Obviously “drilling down” through these “layers”, as Anderson puts it, requires the creation of a multi-pronged linking strategy for automatically disambiguating the semantic complexity that lies between the word in context and the termbase side of the lex/term pair. The notion of such a series of links presupposes that the existing termbase possesses the necessary data modeling features and the explicit information required to forge the links. The likelihood that it does is questionable considering the fact that most existing termbases, even good ones, have been designed for analysis by human translators. As noted above, most of the semantic and transfer detail provided in the termbase is present in the form of non-parsable data that are not readily accessible to

Figure 2: Sager's Link Model



(Where "subject" designates a subject *field* or domain)

automatic processing routines. In fact, Ahmad et al. (1996:170) in their analysis of these issues question whether existing termbase resources are at all appropriate to the kind of cross-application leveraging described in Figure 1.

4 Planning for Multifunctional Lex-Term Resources

4.1 Sager's Theoretical Model

In recommending the creation of multipurpose databases that could support either application-independent lexbase or termbase views of the data, Sager (1990:172) proposed a theoretical model for lexical data that allows for the full documentation of both word- and concept-oriented information, as well as for the integration of all manner of language technology approaches (traditional lexbases and termbases, thesauri, classification systems, ontologies, controlled languages, etc.), a broad intention that is echoed in Ahmad et al. Sager visualizes a network of relational links designed to facilitate optimum flexibility in accessing the data from different views. Most importantly, he recognizes the many-to-many relationships required of such a system and proposes relational models for generating subject field-term-concept links and language-term-usage-concept strategies designed to document the semantic information required to associate a lexical unit with its associated sense. The tables that represent these relationships provide the links that can be exploited to access the data from a number of different perspectives. The argument can be made, however, that the constant regeneration of these links in multiple complex tables may create excessive computational overhead for practical application.

4.2 Current Implementations

There have been a number of projects, both in the public and the private sector, designed either to create a planned lex/term-base or to provide an interchange environment for utilizing both lexbase and termbase data from existing resources. The latter scenario could involve leveraging data in one type of database for incorporation into the other, i.e., the mining of information in order to reuse existing knowledge in a new environment as sketched out in Figure 1.

4.3 TRANSTERM

Whereas Sager's description of his ideal lex/term-base structure seeks to unify the lexicographical and terminological elements in the database into a seamless whole, some developers who have created these kinds of hybrid systems work with a double-sided entry structure similar to the one represented in Figure 1. In their description of the TRANSTERM project, Ahmad et al. outline their approach as possessing multiple "layers" (not the same layers as those outlined by Anderson), one with "terminology specific data" and the other with "language specific data". The latter comprises the semantic, syntactic, and morphological data typically required of lexbase resources such as those created for MT or other NLP tools. The linking element between the two sides of the entry is characterized as a "conduit between the terminology-specific data and the language-specific data" (Ahmad et al.:168) and is called the "TransTerm Linguistic Representative". Just what this element is, the 1996 article does not say, since unfortunately the reports from the TRANSTERM project are not a matter of public record. It could be a form of the word or term itself, a set of relational tables similar to Sager's model, or perhaps simply a pointer linking the two components. Its exact form may be simple, for the more complex elements required of a multi-pronged key are distributed among Ahmad's six supplemental "layers": the terminological container, concept, and unit on one side, and the semantic, syntactic, and morphological units on the other. It is here that Anderson's pragmatic concerns could also be reflected in a way that would make them more accessible for MT than would be the case if they were included in an unparsable transfer comment on the termbase side of the equation.

4.4 TransLexis

IBM's TransLexis program combines the capability of a machine translation lexicon with the kind of terminology management tool required by human technical writers and translators (Schmidt et al. 1999; TransLexis). The program is designed for integration with IBM's proprietary translation memory and MT systems and has been configured for interaction with the Logos MT system. The pivotal elements that appear to hold the lexicographical and terminological information together are simple: the lemma (headword), a language [code?], and a part of speech. Subject field data are presumably important, as is information on homonyms and polysemy, but the first three items cited here are the only obligatory ones.

4.5 OTELO and OLIF

Whereas TRANSTERM and TransLexis are lex/term-base systems that were designed with multifunctionality in mind, the purpose behind OLIF (Open Lexicon Interchange Format) is to facilitate access to both termbase and lexbase data residing in multiple systems and involving different types of resources (Thurmair et al. 1999). A component of the OTELO project (Open Translation Environment for Localization), OLIF was primarily intended to exchange lexical resources among different translation tools, especially in conjunction with OTELO's Common Lexical Database (CLDB).

The OTELO CLDB comes to terms with complexity by creating separate containers for what is called the *linguistic description* and the *terminological description*. A *central entry* (CS) consisting of the canonical form of a term (lemma), a syntactic category (part of speech), a domain classification, and a language code provides the bridge element that links these elements with each other or with other entries. The potential for complexity with respect to basic entry structure is controlled by reducing the individual entry or sub-entry to one term (headword) and one concept in one language. All one-to-many relationships on either the lexbase (polysemy) or termbase (synonyms and equivalents) sides of the composite entry are facilitated by relational links: *cross-reference links* (XREF) for monolingual links and *transfer links* (XFR) for cross-language equivalents. Example 1, p. 17 shows a short sample OLIF entry taken from a Logos™ lexbase.

4.6 MARTIF and X-MARTIF

The MARTIF standard (ISO 12200:1999) as a stand-alone format was designed strictly for the interchange of terminological data. In conjunction with its companion data category standard, ISO 12620:1999, it accommodates the total range of strictly terminological termbase systems from the simplest to the most complex, and covering both prescriptive and descriptive terminology management. This apparent "failure" to include the lexicographical approach is both practical and programmatic. The origins of MARTIF lie in the Text Encoding Initiative, which had its own separate dictionary and NLP lexicon groups (Melby et al. 1994). Unfortunately, original intentions to maintain close coordination among the groups failed when the NLP group failed to produce a final chapter for the guidelines at that time. Furthermore, the creators of MARTIF are for the most part exclusively involved in terminology activities and the creation of TMS applications for human translators.

Current prospects for interchange of information within integrated environments are much more promising, however. MARTIF has been accepted as an international standard and many program developers are pressing to complete their MARTIF implementations. In the meantime, LISA's OSCAR has approved the TMX (Translation Memory eXchange) standard, and OTELO's OLIF format, as demonstrated above, provides a medium for exchanging and merging data to facilitate broader multifunctional solutions. The new Draft Amendment proposal submitted to ISO TC 37 suggests, along with accommodations for the fast-emerging XML standard (eXtensible Markup Language, see XML 1999), that adaptations be made to coordinate MARTIF with TMX and OLIF (ISO/TC 37 /SC 3 N 318). This approach for Extended MARTIF (X-MARTIF) does not

envision that any of these standards will absorb the other(s), but rather that minor additions and modifications will be implemented to allow seamless mutual operation.

One major argument in favor of coordinating these three standards (X-MARTIF, TMX, and OLIF) is that they are all modular standards that incorporate simple, straight-forward models for each of their components. This inherent structural simplicity makes it easier to communicate among the respective systems and above all else makes them attractive to major industrial interests, such as the software localizers in the OSCAR group.

Because OLIF already has encapsulated linguistic data in a separate container, and other elements of the OLIF entry are already covered by basic MARTIF, the extra lexbase data can be provided for by introducing two new major generic identifiers (GIs) to the MARTIF Document Type Definition (DTD): <NLPDesc> and <NLPSet>, which together with a short set of other secondary identifiers accommodate the principle components of the OLIF format.

<NLPDesc> is designed for use in the X-MARTIF header to the NLP features that are allowed in the current document, and the <NLPSet> is used to list a set of feature-value pairs that would apply to a concept entry. The net effect on the original MARTIF DTD of these additions is as follows:
(For basic information on ISO 12200, check the standard itself and BYU/TRG 1999.)

In ISO 12200, Figure 9, Section 4, the statement

<!ELEMENT (sourceDesc|encodingDesc|change) (p+) >

becomes

```
<!ELEMENT (sourceDesc|change) (p+) >
<!ELEMENT encodingDesc (ude?, NLPDesc?, p+) >
<!ELEMENT ude (map+) >
<!ELEMENT map EMPTY >
…
<!ELEMENT NLPDesc (p*, feaDef+) >   <!-- for header information -->
<!ELEMENT NLPfeaDef (feaName, feaValue, feaComment?) >
<!ELEMENT NLPfeaName (#PCDATA) >
<!ELEMENT NLPfeaValue (#PCDATA) >
```

(NOTE: The map element is used when needed to accommodate characters that are not available in the UNICODE character set.)

In Figure 10, the definition of ntig becomes (termGrp, %auxInfo;, NLPSet).

```
<!ELEMENT NLPSet (NLPMono, NLPAllo?, ref*) > <!-- for NLP features on a
term -->
<!ELEMENT NLPMono (NLPfea*) >
<!ELEMENT NLPFea (#PCDATA) >
<!ATTLIST NLPFea feaName CDATA #REQUIRED >
```

<!ELEMENT NLPAllo (NLPForm, NLPInfl) >
<!ELEMENT NLPForm (#PCDATA) >
<!ELEMENT NLPInfl (#PCDATA) >

The names are mnemonic: NLPfeaDef is a feature definition; NLPfeaName is a feature name, etc. The <NLPMONO> section of an <NLPSet> element is the NLP information about a particular term in a particular language (i.e., a monolingual subentry), and the <NLPAllo> element documents allomorphs of a base form. (Example 1 does not contain any allomorphic information, but we will try post an example to BYU/TRG.)

If the optional ude and NLPDesc elements are not used in a MARTIF document, it conforms to the original MARTIF DTD. If one chooses not to use the NLP extensions, NLP features can be represented in the original DTD using custom data category names.

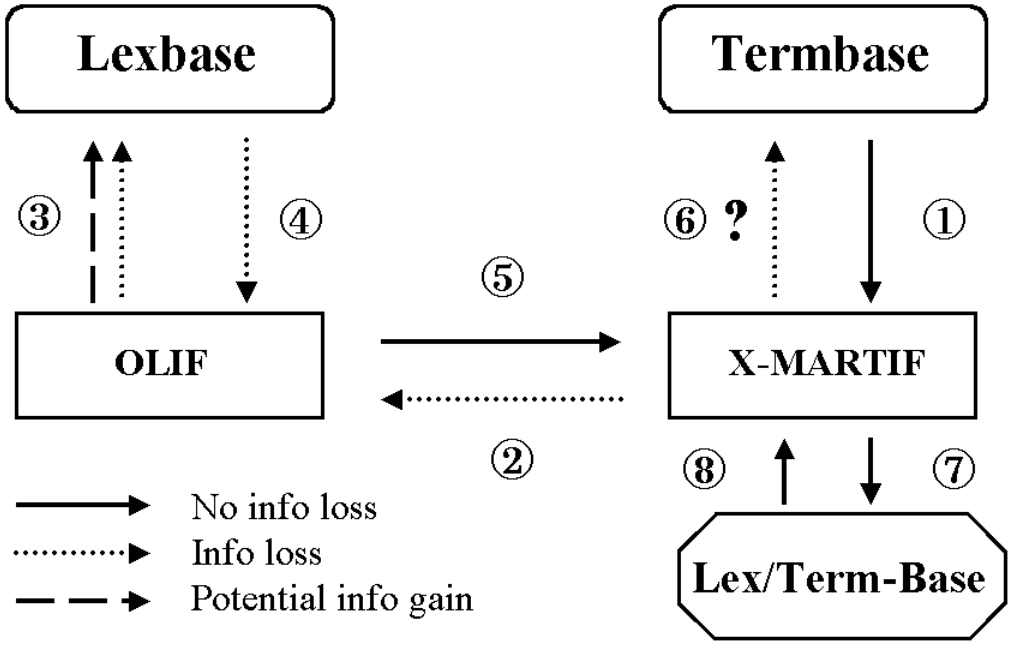
Example 2, p. 19 illustrates a complete termbase entry designed to document the same concept treated in Example 1. Unfortunately, there is insufficient room to reproduce the entire entry in this article, but the complete example is available at BYU/TRG. Anyone familiar with the MARTIF DTD may be startled by the presence of an <admin> element inside both a <termGrp> and a <descripGrp>. The original MARTIF format preferenced the creation of bibliographic entries as shared references residing in front or back matter, but the OSCAR group has requested this accommodation for the inclusion of inline bibliographic references as well. This makes good sense, especially with the growing prevalence of URLs and other potentially one-of-a-kind resources used to document termbase elements.

5 A Data-Leveraging Model for a Multifunctional Information Sharing Environment

Figure 3 illustrates an information-sharing environment where data could flow throughout a multifunctional system designed to leverage the information contained in the different applications for use in other environments. Any data-exchange transaction could be initiated at any point in this sequence. We have only designated the flow from a termbase to extended MARTIF (X-MARTIF) as operation ① because we have to start somewhere. Another factor that has to be taken into consideration in looking at this model is whether the lexbase and termbase involved are always the same known databases or whether they could represent several different databases at different times. In the former case, we can speak of a "roundtrip" with respect to the flow of data through the system, and the absolute ideal objective might be to achieve a situation where a "lossless roundtrip" might be possible, but we will see that this is not highly likely. We will try to keep in mind both scenarios, i.e., the same databases vs. differing databases, as we move along through the model. The representation of the arrows with respect to information loss (or potential gain!) constitutes the primary objective of the diagram.

① MARTIF as it now stands is designed to receive termbase information without loss of content or hierarchical structure. Since X-MARTIF contains MARTIF as a wholly included subset, this step of the process should not require further discussion.

Figure 3: A Data Leveraging Model



② The conversion from X-MARTIF to OLIF should not represent significant content loss. Although MARTIF features a range of data categories not normally treated in OLIF entries, OLIF provides the option of declaring new elements when necessary, so it can be hoped that information loss due to actual data category loss per se could be kept to a minimum. The question always arises, however, whether it is desirable or useful to retain all elements of information, in which case a conscious choice can be made to jettison elements that are potentially irrelevant in the exchange environment. What will inevitably be lost, however, is the detailed embedding structure contained in the X-MARTIF entries that cannot be successfully represented in OLIF. Furthermore, MARTIF features explicit SGML-style (ID-IDRef) links, which can become potentially ambiguous links in OLIF. It may not be possible to reconstitute this information later if and when the data flows back to X-MARTIF after having passed through OLIF and possibly merged with lexbase data.

③ The information loss represented between OLIF and the lexbase is not necessarily unavoidable or even undesirable. It reflects the differences in intent and design for individual lexbases and will vary depending on the lexbase. One can assume that in most cases, strictly terminological information such as detailed definitions, contexts, and other textual material would be stripped away. Of course, this “lexbase” could also be a lex/term-base itself such as TransLexis, in which case more terminological information would be retained.

The potential for information *gain* at this point in the process may look puzzling on first consideration. It is important to bear in mind that as information that originates from a termbase makes its way through the model, it will arrive at the lexbase without some of the critical information needed by the lexbase: i.e., the semantic and syntactical linguistic data that characterizes NLP-oriented lexbases. But as Thurmair et al. point out (1999:238), *some of the high-end machine translation systems have powerful defaulting mechanisms to create complete entries from only partial information*. Bearing in mind that errors can occur at this stage, in many cases upon entering the lexbase, the data might now include value-added information, some of which could potentially then pass on through the system without loss.

Not shown in Figure 3 for lack of a comfortable place for it in the drawing is the exchange of data between OLIF and the OTELO lexical database (CLDB). Although this is, of course, also an extremely important concern, it is a well-documented step that should involve no relevant data loss.

④ The loss of data indicated during operation ④ reflects the loss of hierarchical information from the individual lexbase that is not accommodated by OLIF. This loss is a programmatic one designed to compensate for the tremendous variation among the various databases that may use OLIF as an interchange format. Since this data is not actually usable in other lexbases, or in our scenario, in termbase systems, this loss should be viewed as a system feature more than as a deficiency. It should also be noted that even as the termbase material lacks important information needed for the lexbase, any data that originates at point ④ and starts off in the direction of the termbase will probably not contain information needed for complete terminological documentation, but it can provide the essential information for starting a truncated terminological entry.

⑤ When the data returns to or arrives at X-MARTIF, there should be no additional data loss because amended X-MARTIF is designed to fully contain all information included in OLIF. Ideally, if this process does indeed represent a circular flow (as opposed to situations where the exchange scenario would begin on the lexbase side of the equation), the system could be configured to recognize any data loss that occurred in the original X-MARTIF OLIF conversion and to reconstitute that information upon its return to X-MARTIF.

The question mark here reflects uncertainty concerning the termbase involved in the return exchange. Of course, any linguistic information derived from the lexbase for which there are no “slots” in the termbase will be lost. In the case of a roundtrip, if the data are flowing back into the same starter termbase, there should be no loss of terminological information. However, if data are being imported via X-MARTIF into a different database, any potential loss depends on the relative richness of the termbases in question. As illustrated by Reinke and Schmitz (1998), data loss between termbases will always occur if data are exported from a rich format to a lean format.

The lex/term-base represented in the lower right hand corner of the drawing differs from the CBLD in that it retains all the terminological information contained in the original (or the reconstituted) X-MARTIF file, whereas the CBLD data will reflect the data loss documented during operations ②, ③, and ④. X-MARTIF and the lex/term-base share the same data model, the difference between them being that X-MARTIF is an interchange format and the lex/term-base will be a fully functional relational database implemented using a modified version of the RelTef model (BYU/TRG 1999). Thus the flow of data back and forth between these two components should be lossless.

As was the case between OLIF and the CBLD, data flow from the theoretical lex/term-base back to X-MARTIF should be uncompromised by information loss.

6 Feasibility and Return on Investment (ROI)

Ahmad et al. (1999) make an impassioned plea for the creation of "linguistically informed" composite resources in order to enable later leveraging of such data for NLP purposes, even if current project-related demands do not seem to dictate the need for recording complex linguistic data or coding it in machine-interpretable form. This argument parallels Galinski's earlier plaidoyer for incorporating additional multi-purpose terminological detail during the early stages of lexbase development in anticipation of broader resource exploitation later in the life cycle of the project (Galinski 1988:10). Whether the goal is increased linguistic information (Ahmad) or additional terminological detail (Galinski), the primary deterrent to realizing these objectives is the fact that in most commercial environments, system designers strive to achieve maximum immediate payback with minimum data collection costs. Unless there is a viable prospect that the additional effort will eventually pay off in real return on investment, it is highly unlikely that this potentially expensive approach will be adopted merely for the sake of satisfying as-yet undefined needs. Nevertheless, as more and more enterprises adopt TM and MT systems in conjunction with TMSs, the value of the more complex approach is becoming apparent to more people.

In the absence of increased investment in either linguistic data appended to terminological entries or terminological information included in lexicographical resources, the question remains whether there is value to be gained by attempting to leverage the information that is indeed included in collections that were not originally designed for multi-purpose application. Ahmad is certainly correct in noting that the absence of semantic and syntactic information impedes direct, immediate automatic importation or exploitation of information, but Thurmair's experience with the "powerful defaulting mechanisms" implies that at least some of the required information can be generated automatically. Especially when dealing with terminological resources whose high quality merits the attempt to reuse the data they contain, it certainly seems feasible to pursue this effort, even it the constraints imposed by these issues dictate that only *some* of the required information can be loaded without (potentially costly) human intervention. With careful planning, this still seems more efficient than starting from scratch.

Bibliography

AHMAD, K. OGONOWSKI, A.; DAUPHIN, E.; STA, J.-D.; and ARPPE, A. (1996): Engineering Terminology – A Case for a Linguistically-informed Terminology Database. TKE ’96: Terminology and Knowledge Engineering. Frankfurt am Main: Indeks Verlag.

ANDERSON, R. J. (1999): Factors Influencing Threaded Fasteners in Industrial Assembly. Unpublished M.A. thesis. Kent: Kent State University Institute for Applied Linguistics.

BUDIN, G. (1996): Wissensorganisation und Terminologie: Die Komplexität und Dynamik wissenschaftlicher Informations- und Kommunikationsprozesse. Tübingen: Gunter Narr Verlag.

BYU/TRG. (1999):http://www.ttt.org. Accessed 1999-04-11.

CARROLL, S. (1999): The Technical content of TMX 1.1: A format for the exchange of data between competing translation database systems. Multilingual Computing & Technology. No. 22, Vol. 9/6, 48-49.

FELBER, H. (1984): Terminology Manual. Paris: UNESCO and Infoterm.

GALINSKI, C. 1988. Advanced Terminology Banks Supporting Knowledge-Based MT: Some reflections on the costs for setting up and operating a terminological data bank (TDB). Conference Proceedings of New Directions in Machine Translation. Dordrecht/Providence: Foris Publishers, 1-15.

ISO 12200:1999, Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) – Negotiated interchange. Geneva: ISO.

ISO 12620:1999, Computer applications in terminology – Data categories. Geneva: ISO.

ISO/TC 37 /SC 3 N 318: ISO 12200:1999. Computer applications in terminology – Machine readable terminology interchange format (MARTIF) – Negotiated Interchange – Amendment 1: (to include a normative annex H): Extended MARTIF. ISO TC 37/SC 3 working document.

MELBY, A.K. (1982): Multi-level Translation Aids in a Distributed System In: North Holland Linguistic Series # 47 — Proceedings of COLING ’82. Jan Horecky, ed. Amsterdam: North Holland Publishing Company, 215-220.

MELBY, A. K. (1988): Electronic Dictionaries and the Translator’s Workbench. Manuscripts and Program of the International Symposium on Electronic Dictionaries. Tokyo: ISED, 62-64.

MELBY, A. K.; BUDIN, G.; SHREVE, G. M. and WRIGHT, S. E. (1994): Base Tag Set for Terminological Databases. In: Guidelines for Electronic Text Encoding and Interchange, TEI P3, C.M. Sperberg-McQueen and L. Burnard (eds). Chicago and Oxford: ACH, ACL, ALLC, 371-390.

OSCAR. (1999):html://www.lisa.unige.ch/tmx/index.html. Accessed 1999-04-04.

REINKE, U. and SCHMITZ, K.-D. (1998): Testing the Machine Readable Terminology Interchange Format (MARTIF). Saarbrücker Studien zu Sprachdatenverarbeitung und Übersetzen. Saarbrücken: Fachrichtung 8.6–Angewandte Sprachwissenschaft sowie Übersetzen und Dolmetschen –Universität des Saarlandes.

RIGGS, F. W., Ed. (1985): Ethnicity – INTERCOCTA Glossary: Concepts and Terms used in Ethnicity Research. International Conceptual Encyclopedia for the Social Sciences: Vol. 1. Honolulu: INTERCOCTA**.**

SAGER, J.C. (1990): A Practical Course in Terminology Processing. Amsterdam: John Benjamins Publishing Company.

SCHMIDT, G.; HEJL, M; JORNITZ, G.; LUEDDE, A.; PAINKE, E.; and SCHOECK, E. (1999): Overview of Current IBM Translation Technology. TAMA ’98–Terminology in Advanced Microcomputer Applications. Proceedings of the 4th TermNet Symposium: Tools for Multilingual Communication. Vienna: TermNet, 25-39.

THURMAIR, G.; RITZKE, J.; and MC CORMICK, S. (1999): The Open Lexicon Interchange Format –OLIF. TAMA '98–Terminology in Advanced Microcomputer Applications. Proceedings of the 4th TermNet Symposium: Tools for Multilingual Communication. Vienna: TermNet. 237-262.

TransLexis. (1999):http://www.qsoft.de/ibmtrans/translex.htm. Accessed 1999-04-01.
XML. (1999):http://msdn.microsoft.com/xml.default. Accessed 1999-04-11.

WRIGHT, S. E., and BUDIN, G. (1997): The Handbook of Terminology Management, Vol. I. Amsterdam and Philadelphia: John Benjamins Publishing Company.

WÜSTER, E. (1973/1985): Einführung in die allgemeine Terminologielehre und terminologische Lexikographie. 2nd Edition. Copenhagen: Handelshochschule Kopenhagen.

The authors wish to acknowledge the assistance of Logos Corporation and the OTELO group for providing support materials used in this article.

Example 1: OLIF lex-termbase entry

The following text should be read as newspaper columns. Items in boldface are explained in the second table.

<OLIF>
<HEADER>
<AUTHOR = logos>
<DATE = 1999-05-04>
<CHARACTER CODE = ISO LATIN 8859/1>
<PROJECT = mt>
<SOURCE = logos>
<TARGET = otelo>
</HEADER>

<BODY>

<ENTRY>
<MONO>
<CAN = Mutterschaftsgeld>
<CAT = noun>
<SA = log-01-000>
<LG = de>
<TSTAT = mt>
<CE-CMT = transfer of 'maternity benefit' [en]>
<CE-AUTHOR = logos>
<CE-DATE = 1997-04-12>
<L-DATE = 1997-04-12>
<GD = (n)>
<USE = offline>
</MONO>
</ENTRY>

<ENTRY>
<MONO>
<CAN = assegno di maternità>
<CAT = noun>
<SA = log-01-000>
<LG = it>
<ETYP = mw>
<TSTAT = mt>
<CE-CMT = transfer of 'maternity benefit' [en]>
<CE-AUTHOR = logos>
<CE-DATE = 1997-04-12>
<L-DATE = 1997-04-12>
<GD = (m)>
<USE = offline>
</MONO>
</ENTRY>

<ENTRY>
<MONO>
<CAN = allocation de maternité>
<CAT = noun>
<SA = log-01-000>
<LG = fr>
<ETYP = mw>
<TSTAT = mt>
<CE-CMT = transfer of 'maternity benefit' [en]>
<CE-AUTHOR = logos>
<CE-DATE = 1997-04-12>
<L-DATE = 1997-04-12>
<GD = (f)>
<USE = offline>
</MONO>
</ENTRY>

<ENTRY>
<MONO>
<CAN = maternity benefit>
<CAT = noun>
<SA = log-01-000>
<LG = en>
<ETYP = mw>
<TSTAT = mt>
<CE-AUTHOR = logos>
<CE-DATE = 1997-04-12>
<L-DATE = 1997-04-12>
<SEMT = (cnt,mea)>
<SYNT = (cnt)>
<USE = offline>
</MONO>

<XFR>
<CAN = Mutterschaftsgeld>
<CAT = noun>
<SA = log-01-000>
<LG = de>
<EQ = sub>
<X-SRC = logos>
<AUTHOR = logos>
<DATE = 1997-04-12>
</XFR>

<XFR>
<CAN = allocation de maternité>
<CAT = noun>
<SA = log-01-000>
<LG = fr>
<EQ = sub>
<X-SRC = logos>
<AUTHOR = logos>
<DATE = 1997-04-12>
</XFR>
<XFR>
<CAN = beneficio por maternidad>
<CAT = noun>
<SA = log-01-000>
<LG = es>
<EQ = sub>
<X-SRC = logos>
<AUTHOR = logos>
<DATE = 1997-04-12>
</XFR>
<XFR>
<CAN = assegno di maternità>
<CAT = noun>
<SA = log-01-000>
<LG = it>
<EQ = sub>
<X-SRC = logos>
<AUTHOR = logos>
<DATE = 1997-04-12>
</XFR>
</ENTRY>

<ENTRY>
<MONO>
<CAN = beneficio de maternidad>
<CAT = noun>
<SA = log-01-000>
<LG = es>
<ETYP = mw>
<TSTAT = mt>
<CE-CMT = transfer of 'maternity benefit' [en]>
<CE-AUTHOR = logos>
<CE-DATE = 1997-04-12>
<L-DATE = 1997-04-12>
<SNU = plt>
<GD = (m)>
<USE = offline>
</MONO>
</ENTRY>
</BODY>
</OLIF>

The following table provides explanatory information for the boldfaced OLIF data categories. The columns should be read as parallel columns.

<OLIF> <HEADER> <AUTHOR = logos> <DATE = 1999-05-04> <CHARACTER CODE = ISO LATIN 8859/1> <PROJECT = mt> <SOURCE = logos> <TARGET = otelo> </HEADER>	These data items constitute NLPDesc information that will be included in the header of the X-MARTIF file.
<ENTRY> <MONO> <CAN = Mutterschaftsgeld> <CAT = noun> <SA = log-01-000>	Since the <ENTRY> <MONO> set treats a single term in a single language, this information is subsumed into an <ntig> within an X-MARTIF <termEntry>. Canonical form of a term; <term> <termNote type='partOfSpeech'>noun… Subject area; retain inside the <NLPSet> to ensure low-loss roundtrip unless the subject

<LG = de> <TSTAT = mt> <CE-CMT = transfer of 'maternity benefit' [en]> <CE-AUTHOR = logos> <CE-DATE = 1997-04-12> <L-DATE = 1997-04-12> <GD = (n)> <USE = offline>	field set of the X-MARTIF file has been harmonized with that of the OLIF file <langSet lang='de'> Technical state [status]: MT only; retain in NLPSet Central Entry-Comment; implicit in this note is the fact that English serves as the pivot language, which in turn implies directionality throughout the concept-oriented set of data entries; retain in <NLPSet> Administrative information which can simply be retained in the <NLPSet> on the assumption that this info may differ from the parallel information for the rest of the X-MARTIF entry. <termNote type='gender'> Retain in <NLPSet>
<ETYP = mw>	Entry type = multiword term; retain
<SEMT = (cnt,mea)> <SYNT = (cnt)>	Semantic type = countable [count noun], measure; retain Syntactic type=countable; retain
<SNU = plt>	Special number=pluraletantum; retain
<XFR> <EQ = sub> <X-SRC = logos> <AUTHOR = logos> <DATE = 1997-04-12>	Transfer entry, foreign language equivalent This relationship is implicit by virtue of the fact that the terms in the different languages reside in the same entry in the X-MARTIF entry. Retain in <NLPSet> or: <termNote type='degreeOfEquivalence'>narrower Identical to info from <mono> entry; reconstitute during re-importation to OLIF

Example 2: MARTIF Representation of a Full Termbase Entry + NLPSet

```
<?XML version="1.0"?>
<!DOCTYPE martif PUBLIC "ISO 12200:1999//DTD MARTIF core (Part 2
V1)//EN" [
<!ENTITY % datcats PUBLIC 'ISO 12200:1999//DTD MARTIF datcats Maxi-set
(Part 2 V1)//EN'>
]>

<martif type=Part2v1 lang=en>

<martifHeader>

<fileDesc>
<sourceDesc><p>sample X-MARTIF entry from TKE99 paper</p></sourceDesc>
</fileDesc>

<encodingDesc>
```

maxi-set
1
</encodingDesc>

</martifHeader>

<text>
<body>

<termEntry id='C007'>
<descrip type='subjectField'>insurance</descrip>
<descrip type='definition'>Financial support paid to a new mother.</descrip>
<adminGrp><admin type='responsibility'>super</admin>
<date type='origination'>99/04/08 - 09:34:51</date></adminGrp>

<langSet lang=en>
<ntig>
<termGrp><term>maternity benefit</term>
<termNote type='partOfSpeech'>noun</termNote>
<termNote type='geographicalUsage'>GB</termNote>
<termNote type='transferComment'>Note that the British maternity benefit is a one-time payment, in contrast to the payments made to mothers under the law of some other countries.</termNote>
<admin type='source'>http://www.actuaries.ca/lexicon/lextm.html'</admin></termGrp>

<descripGrp><descrip type='definition'>A once-only payment (currently £100) from the Social Fund for each baby expected, born or adopted.</descrip>
<admin type='source'>http://www.babyworld.com/bp00009.htm#maternity absence</admin>
</descripGrp>
<descripGrp><descrip type='context'>The maternity benefit is a once-only payment (currently £100) from the Social Fund for each baby expected, born or adopted. Payment (which does not depend on National Insurance contributions) is available only where a mother and/or her partner receives Income Support, Family Credit or Disability Working Allowance.</descrip>
<admin type='source'>http://www.babyworld.com/bp00009.htm#maternity absence</admin>
</descripGrp>
<NLPSet>
<NLPMono>
<NLPFea feaName='SA'>log-01-000</NLPFea>
<NLPFea feaName='ETYP'>mw</NLPFea>
<NLPFea feaName='TSTAT'>mt</NLPFea>
<NLPFea feaName='CE-AUTHOR'>logos</NLPFea>
<NLPFea feaName='CE-DATE'>1997-04-12</NLPFea>
<NLPFea feaName='L-DATE'>1997-04-12</NLPFea>
<NLPFea feaName='SEMT'>(cnt,mea)</NLPFea>
<NLPFea feaName='SYNT'>(cnt)</NLPFea>
<NLPFea feaName='USE'>offline</NLPFea>
</NLPMono>
</NLPSet>
</ntig>
</langSet>

<langSet lang='de'>
<ntig >
<termGrp><term>Mutterschaftsgeld</term>
<termNote type='partOfSpeech'>noun</termNote>
<termNote type='grammaticalGender'>neuter</termNote></termGrp>

<descripGrp><descrip type='definition'>Die der Mutter für die Dauer des Beschäftigungsverbotes nach der Entbindung aufgrund des Mutterschutzgesetzes (acht Wochen bzw. zwölf Wochen bei Früh- und Mehrlingsgeburten) zustehende Geldzahlung.</descrip>
<admin type='source'>http://www.bawue.gew.de/fundus/erziehgeld.html</admin></descripGrp>

<ref type='superordinateConceptGeneric' target='id008'>Leistung bei Mutterschaft</ref>
<NLPSet>
<NLPMono>
<NLPFea feaName='SA'>log-01-000</NLPFea>
<NLPFea feaName='TSTAT'>mt</NLPFea>
<NLPFea feaName='CE-CMT'>transfer of 'maternity benefit' [en]</NLPFea>
<NLPFea feaName='CE-AUTHOR'>logos</NLPFea>
<NLPFea feaName='CE-DATE'>1997-04-12</NLPFea>
<NLPFea feaName='L-DATE'>1997-04-12</NLPFea>
<NLPFea feaName='USE'>offline</NLPFea>
</NLPMono>
</NLPSet>
</ntig>
</langSet>

<langSet lang='fr'>
<ntig>
<termGrp><term>prestation de maternité</term>
<termNote type='partOfSpeech'>noun</termNote>
<termNote type='grammaticalGender'>feminine</termNote>
<termNote type='geographicalUsage'>CA</termNote>
<admin type='source'>http://www.actuaries.ca/lexicon/lextm.html</admin></termGrp>

<descripGrp><descrip type='context'>Régime de prestations supplémentaires de chômage: pour chacune des semaines où elle reçoit ou pourrait recevoir des prestations d'assurance-chômage, une indemnité complémentaire égale à la différence entre quatre-vingt-dix pour cent (90%) de son salaire hebdomadaire et la prestation d'assurance-chômage qu'elle reçoit ou pourrait recevoir sans tenir compte de toute réduction du nombre de semaines pendant lesquelles elle bénéficie de prestation de maternité, mais sans toutefois excéder quinze (15) semaines...</descrip>

<admin
type='source'>http://www.ville.montreal.qc.ca/personnel/conventions/bleus/articles/art31.htm

</admin></descripGrp>
</ntig>

<ntig>
<termGrp><term>allocation de maternité</term>
<termNote type='partOfSpeech'>noun</termNote>
<termNote type='grammaticalGender'>feminine</termNote>
<termNote type='geographicalUsage'>CA</termNote>
<admin type='source'>http://www.actuaries.ca/lexicon/lextm.html</admin></termGrp>

<descripGrp><descrip type='definition'>Une compensation financière à la travailleuse salariée qui doit s'absenter du travail pour cause de grossesse et qui satisfait aux critères d'admissibilité.</descrip>
<admin
type='source'>http://www.gouv.qc.ca/social/famille.htm#maternite</admin></descripGrp>

<descripGrp><descrip type='context'>Le <hi type='entailedTerm' target='id009'>Programme d'allocation de maternité</hi> (<hi type='entailedTerm' target='id009'>PRALMA</hi>) apporte une

compensation financière à la travailleuse salariée qui doit s'absenter du travail pour cause de grossesse et qui satisfait aux critères d'admissibilité.</descrip>
<admin type='source'>http://www.gouv.qc.ca/social/famille.htm#maternite</admin></descripGrp>
<NLPSet>
<NLPMono>
<NLPFea feaName='SA'>log-01-000</NLPFea>
<NLPFea feaName='ETYP'>mw</NLPFea>
<NLPFea feaName='TSTAT'>mt</NLPFea>
<NLPFea feaName='CE-CMT'>transfer of 'maternity benefit' [en]</NLPFea>
<NLPFea feaName='CE-AUTHOR'>logos</NLPFea>
<NLPFea feaName='CE-DATE'>1997-04-12</NLPFea>
<NLPFea feaName='L-DATE'>1997-04-12</NLPFea>
<NLPFea feaName='USE'>offline</NLPFea>
</NLPMono>
</NLPSet>
</ntig>

<ntig>
<termGrp><term>indemnité de maternité</term>
<termNote type='partOfSpeech'>noun</termNote>
<termNote type='grammaticalGender'>feminine</termNote>
<termNote type='geographicalUsage'>CA</termNote>
<admin type='source'>http://www.actuaries.ca/lexicon/lextm.html</admin>
</termGrp>
</ntig>
</langSet>

<langSet lang='es'>
<ntig>
<termGrp><term>beneficio de maternidad</term>
<termNote type='partOfSpeech'>noun</termNote>
<termNote type='GrammaticalGender'>masculine</termNote>
<termNote type='geographicalUsage'>Latin America</termNote>
<termNote type='normativeAuthorization'>preferred term</termNote>
<termNote type='transferComment'>Contexts found for this term in Spanish pertain primarily to health benefits, but there is no reason to assume that the term would be incorrect if used as an equivalent for the terms pertaining to financial benefits paid to the mother.</termNote>
<admin type='source'>http://www.minerva.com.ar/reglam3.html</admin></termGrp>
<descripGrp><descrip type='context'>Sólo podrá obtener el servicio y/o beneficio de maternidad, la titular afiliada a la categoría que contempla dicho beneficio, o la cónyuge del titular que figure como tal en la solicitud de ingreso, siempre que hayan cumplido la carencia correspondiente. Otras beneficiarias incluidas en la solicitud de ingreso (hija, familiar a cargo), no gozan de la cobertura de maternidad.</descrip>
<admin type='source'>http://www.minerva.com.ar/reglam3.html</admin></descripGrp>

<descripGrp><descrip type='context'>Beneficio de maternidad y recién nacido</descrip>
<admin type='source'>http://www.ssspr.com/servmat.htm</admin></descripGrp>
<NLPSet>
<NLPMono>
<NLPFea feaName='SA'>log-01-000</NLPFea>
<NLPFea feaName='ETYP'>mw</NLPFea>
<NLPFea feaName='TSTAT'>mt</NLPFea>
<NLPFea feaName='CE-CMT'>transfer of 'maternity benefit' [en]</NLPFea>
<NLPFea feaName='CE-AUTHOR'>logos</NLPFea>
<NLPFea feaName='CE-DATE'>1997-04-12</NLPFea>
<NLPFea feaName='L-DATE'>1997-04-12</NLPFea>

<NLPFea feaName='SNU'>plt</NLPFea>
<NLPFea feaName='USE'>offline</NLPFea>
</NLPMono>
</NLPSet>
</ntig>
</langSet>

<langSet lang='it'>
<ntig>
<termGrp><term>assegno di maternità</term>
<termNote type='partOfSpeech'>noun</termNote>
<termNote type='grammaticalGender'>masculine</termNote>
<admin type='source'>http://www.elpacifico.com/mvipago.htm</admin></termGrp>
<descripGrp><descrip type='context'>In questo quadro rientrano, ad esempio, le misure dirette ad esentare dall'Irpef la maggiorazione sociale delle pensioni al minimo; l'aumento della detrazione per I pensionati sotto i 18 milioni di reddito; l'aumento delle pensioni sociali e dell'assegno sociale; l'estensione di un assegno di maternità ai soggetti non titolari della specifica tutela riconosciuta, oggi, solo alle lavoratrici dipendenti ed autonome.</descrip>
<admin type='source'>http://web.cisl.it/doc/finanziaria_assegni_familiari.htm</admin>

</descripGrp>
<NLPSet>
<NLPMono>
<NLPFea feaName='SA'>log-01-000</NLPFea>
<NLPFea feaName='ETYP'>mw</NLPFea>
<NLPFea feaName='TSTAT'>mt</NLPFea>
<NLPFea feaName='CE-CMT'>transfer of 'maternity benefit' [en]</NLPFea>
<NLPFea feaName='CE-AUTHOR'>logos</NLPFea>
<NLPFea feaName='CE-DATE'>1997-04-12</NLPFea>
<NLPFea feaName='L-DATE'>1997-04-12</NLPFea>
<NLPFea feaName='USE'>offline</NLPFea>
</NLPMono>
</NLPSet>
</ntig>
</langSet>
</termEntry>

<termEntry id='id008'>
<langSet lang='de'><tig><term>Leistung bei Mutterschaft</term></tig></langSet>
</termEntry>

<termEntry id='id009'>
<langSet lang='de'><tig><term>Programme d'allocation de maternité</term></tig></langSet>
</termEntry>

</body>

</text>

</martif>