From Resources to Applications. Designing the Multilingual ISLE Lexical Entry

Sue Atkins[1], Nuria Bel[2], Francesca Bertagna[3], Pierrette Bouillon[4], Nicoletta Calzolari[5], Christiane Fellbaum[6], Ralph Grishman[7], Alessandro Lenci[3], Catherine MacLeod[7], Martha Palmer[8], Gregor Thurmair[9], Marta Villegas[2], Antonio Zampolli[5]

ITRI, University of Brighton, Lewes Road, Brighton, BN2 4GJ[1]
b.t.s.atkins@itri.brighton.ac.uk

GilcUB, Adolf Florensa s/n, 08028, Barcelona, Spain[2]
{nuria, tona}@gilcub.es

Università di Pisa, Dipartimento di Linguistica - via Santa Maria 36, 56127, Pisa, Italy[3]
{francesca.bertagna, alessandro.lenci}@ilc.cnr.it

ISSCO, Boulevard du Pont-d'Arve, 1211, Genève 4, Switzerland[4]
pierrette.bouillon@issco.unige.ch

Istituto di Linguistica Computazionale – CNR - Area della Ricerca di Pisa, via Moruzzi 1, 56127, Pisa, Italy[5]
{glottolo, pisa}@ilc.cnr.it

Psychology Dept., Princeton University, Princeton, New Jersey, USA[6]
fellbaum@clarity.princeton.edu

New York University, Computer Science Dept., 15 Broadway, 703, New York, USA[7]
{grishman, macleod}@cs.nyu.edu

CIS Dept., University of Pennsylvania, 200 South 33rd Street, Philadelphia, USA[8]
mpalmer@linc.cis.upenn.edu

Sail Labs, Balanstr. 57, 81541, Munich, Germany[9]
gregor.thurmair@sail-labs.de

Abstract

The ISLE Computational Lexicon Working Group is committed to the consensual definition of a standardized infrastructure to develop multilingual resources for HLT applications. In particular, the ISLE-CLWG pursues this goal by designing MILE (*Multilingual ISLE Lexical Entry*), a general schema for the encoding of multilingual lexical information. This has to be intended as a meta-entry, acting as a common representational layer for multilingual lexical resources. We present the general architecture and features of MILE, as well as the methodology adopted for its definition. In particular, we focus on two essential ingredients for the MILE specification: the selection of the types of lexical information most relevant to establish multilingual correspondences, and the specification of a data structure which will provide the formal backbone of the MILE as a general representation language to develop multilingual resources. The ISLE recommendations will also consist of a first repository of *shared lexical objects*, including main syntactic constructions, basic operations and conditions to establish multilingual links, macro-semantic objects, etc., for the encoding of lexical units at a higher level of abstraction, as a step in the direction of simplifying and improving the usability of the MILE recommendations. We are also developing the ISLE Lexicographic tool.

1. Introduction

The ISLE Computational Lexicon Working Group (CLWG) is committed to the consensual definition of a standardized infrastructure to develop multilingual resources for HLT applications, with particular attention to the needs of Machine Translation (MT) and Crosslingual Information Retrieval (CLIR) systems. Compared with other standardization initiatives active in this field (e.g. OLIF-2; cf. Lieske *et al.*, 2001), the original character of ISLE resides in its specifically focusing on the *grey area* of HLT where well-assessed language technology meets more advanced levels and forms of linguistic description.

In particular, various aspects of lexical semantics, although still part of ongoing research, are nevertheless regarded by industrials and developers as the "next-step" in new generation multilingual applications. Standard definition in this area thus means to lay a first bridge between research in multilingual resource development and its exploitation in advanced technological systems. In particular, the ISLE-CLWG pursues this goal by designing MILE (*Multilingual ISLE Lexical Entry*), a general schema for the encoding of multilingual lexical information. This has to be intended as a meta-entry, acting as a common representational layer for multilingual lexical resources.

The aim of this paper is to present the general architecture and features of MILE, as well as to discuss

the methodology adopted for its definition. In section 2 the general objective and methodology adopted by the ISLE CLWG will be illustrated. The general architecture of MILE will be presented in section 3, and in section 4 the current work leading to its definition will be presented. In particular, the paper will focus on the two essential ingredients for the MILE specification: on the one hand the selection of the types of lexical information most relevant to establish mutilingual correspondences, and on the other hand the specification of a data structure which will provide the formal backbone of the MILE as a general representation language to develop multilingual resources and to link computational lexicons.

2. The ISLE Computational Lexicon Working Group

The ISLE[1] (*International Standards for Language Engineering*) project is a continuation of the long standing EAGLES initiative (Calzolari *et al.*, 1996).[2] ISLE is carried out in collaboration between American and European groups in the framework of the EU-US International Research Co-operation, supported by NSF and EC.

EAGLES work towards *de facto* standards has already allowed the field of Language Resources (LR) to establish broad consensus on critical issues for some well-established areas, providing thus a key opportunity for further consolidation and a basis for technological advance. EAGLES previous results have already become *de facto* standards. Existing EAGLES results in the Lexicon and Corpus areas are currently adopted by a number of European - and recently also National – projects (e.g. LE-PAROLE and LE SIMPLE; cf. Ruimy *et al.*, 1998; Lenci *et al.*, 2000), thus becoming "the *de-facto* standard" for LR in Europe.

The current ISLE CLWG aims at selecting mature areas and results in computational lexical semantics and in multilingual lexicons, which can also be regarded as stabilized achievements, thus to be used as the basis for future research. The main objectives of the working group are illustrated below.

2.1. Goals and methodological principles

Lexical semantics has always represented a "*wild frontier*" in the investigation of natural language, let alone when this is also aimed at implementing large-scale systems based on HLT components. In fact, the number of open issues in lexical semantics both on the representational, architectural and content level might induce an actually unjustified negative attitude towards the possibility of designing standards in this difficult territory. Rather to the contrary, standardisation must be conceived as enucleating and singling out - in the open field of lexical semantics - the areas that already present themselves with a clear and high degree of stability, although this is often hidden behind a number of formal differences or representational variants, that prevent the possibility of exploiting and enhancing the aspects of commonality and the already consolidated achievements.

Standards must emerge from state-of-the-art developments. With this respect, the ISLE CLWG adheres to the leading methodological principle that the process of standardization, although by its own nature not intrinsically innovative, *must – and actually does – proceed shoulder to shoulder with the most advanced research.* Consistently, the ISLE standardization process pursues a twofold objective:

i. defining standards both at the content and at the representational level for those aspects of computational lexicons which are already widely used by applications;
ii. proposing recommendations for the areas of computational lexical semantics which are still in the "front line" of ongoing research, but also appear to be ready for their applicative exploitation, and are most required by HLT systems to achieve new technological leap forwards.

This double perspective is one of the peculiar features of the ISLE activities, and contributes to its added value with respect to other current standardization initiatives. This way, ISLE intends on the one hand to answer to the need of fostering the reuse and interchange of existing lexical resources, and on the other hand to enhance the technological transfer from advanced research to applications.

The consolidation of a standards proposal must be viewed, by necessity, as a slow process comprising, after the phase of putting forward proposals, a cyclical phase involving EAGLES external groups and projects with:

- careful evaluation and testing by the scientific community of recommendations in concrete applications;
- application, if appropriate, to a large number of languages;
- feedback on and readjustment of the proposals until a stable platform is reached, upon which a real consensus - acquiring its meaning by real usage - is arrived at;
- dissemination and promotion of consensual proposals.

What can be defined as *new advance* in this process is the highlighting of the areas for consensus (or of the areas in which consensus could be reached) and the gradual consciousness of the stability that evolves within the communities involved. A first benefit is the possibility, for those working in the field, of focusing their attention on as yet unsolved problems without losing time in rediscovering and re-implementing what many others have already worked on. Useful indications of *best practice* will therefore come to researchers as well as resource developers. This is the only way our discipline can really move forward.

[1] ISLE Web Site URL: lingue.ilc.pi.cnr.it/EAGLES96/isle/ISLE_Home_Page.htm

[2] EAGLES stands for *Expert Advisory Group for Language Engineering Standards* and was launched within EC Directorate General XIII's Linguistic Research and Engineering programme in 1993, continued under the Language Engineering programme, and now under the Human Language Technology (HLT) programme as ISLE, since January 2000.

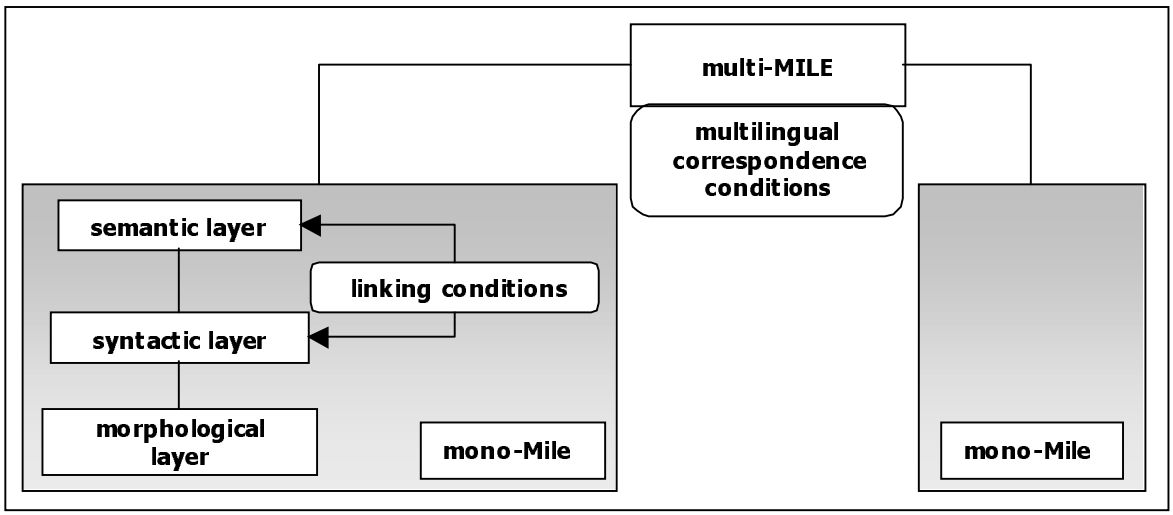

Figure 1: The general architecture of MILE

Finally, one of the targets of standardization, and actually one of the main aims of the ISLE CLWG activities, is to create a common parlance among the various actors (both of the scientific and of the industrial R&D community) in the field of computational lexical semantics and multilingual lexicons, so that synergies will be thus enhanced, commonalities strengthened, and resources and findings usefully shared. In other terms, the process of standard definition undertaken by the CLWG, and by the ISLE enterprise in general, represents an essential interface between advanced research in the field of multilingual lexical semantics, and the practical task of developing resources for HLT systems and applications. It is through this interface that the crucial trade-off between research practice and applicative needs will actually be achieved.

3. The MILE Architecture

In its general design, MILE is envisaged as a highly *modular* and *layered* architecture (see Figure 1), as described in Calzolari *et al.* (2001b). Modularity concerns the "horizontal" MILE organization, in which independent and yet linked modules target different dimensions of lexical entries. On the other hand, at the "vertical" level, a layered organization is necessary to allow for different degrees of granularity of lexical descriptions, so that both "shallow" and "deep" representations of lexical items can be captured. This feature is particularly crucial in order to stay open to the different styles and approaches to the lexicon adopted by existing multilingual systems.

At the top level, MILE includes two main modules, *mono-MILE*, providing monolingual lexical representations, and *multi-MILE*, where multilingual correspondences are defined. With this design choice the ISLE-CLWG intends also to address the particularly complex and yet crucial issue of multilingual resource development through the integration of monolingual computational lexicons. Mono-MILE is organized into independent modules, respectively providing *morphological*, *syntactic* and *semantic* descriptions. The latter surely represents the core and the most challenging part of the ISLE-CLWG activities, together with the two other crucial topics of *collocations* and *multi-word expressions*, which have often remained outside standardization initiatives, and nevertheless have a crucial role at the multilingual level. This bias is motivated by the necessity of providing an answer to the most urgent needs and desiderata of next generation HLT, as also expressed by the industrial partners participating to the project. With respect to the issue of the representation of multi-word expressions in computational lexicons, the ISLE-CLWG is actively cooperating with the NSF sponsored XMELLT project (Calzolari *et al.*, 2002).

Multi-MILE specifies a formal environment for the characterization of multilingual correspondences between lexical items. In particular, source and target lexical entries can be linked by exploiting (possibly combined) aspects of their monolingual descriptions. Moreover, in multi-MILE both syntactic and semantic lexical representations can also be enriched, so as to achieve the granularity of lexical description required to establish proper multilingual correspondences, and which is possibly lacking in the original monolingual lexicons.

According to the ISLE approach, monolingual lexicons can thus be regarded as *pivot lexical repositories*, on top of which various language-to-language multilingual modules can be defined, where lexical correspondences are established by partly exploiting and partly enriching the monolingual descriptions. This architecture guarantees the independence of monolingual descriptions while allowing for the maximum degree of flexibility and consistency in reusing existing monolingual resources to build new bilingual lexicons.

The MILE architecture is intended to provide the common representational environment needed to implement such an approach to multilingual resource development, with the goal of maximizing the reuse, integration and extension of existing monolingual computational lexicons.

4. Building MILE

In the process of specifying the various components of MILE, the ISLE-CLWG has adopted a two-track strategy:

i. identifying the lexical dimensions and the various types of information which are relevant to establish multilingual correspondences;
ii. defining a suitable formal data model to encode this information as well as the operations required at the multilingual level.

The architecture of the PAROLE-SIMPLE lexicons has been selected to provide the necessary bootstrapping basis for the stepwise refinement cycle leading to MILE.

4.1. Multilingual lexical dimensions

To tackle point i) above, the survey of the available computational lexicons and system needs, carried out in the preliminary phases of the project (cf. Calzolari *et al.*, 2001a), has been complemented with a more lexicographic-based effort, to identify the types of information used in bilingual dictionaries to establish translation equivalents. To this purpose, the CLWG has organized two "task forces" with the responsibility respectively of creating a sample of lexical entries and investigating the use of *sense indicators* in traditional bilingual dictionaries. The aim of these activities has been twofold: i) highlighting the various types of information useful to determine the transfer conditions; ii) exploring and evaluating the full expressive potentialities provided by the reference computational model (i.e. the PAROLE-SIMPLE architecture).

4.1.1. Sample lexical entries

A number of lexical items (nouns, verbs and adjectives) have been selected on the basis of their degree of polysemy and complexity of translation, to build a general "test suite" of possible multilingual transfer scenarios. This experiment has started with English, Italian, French and German lexical entries, with the goal of extending it to other languages (such as for instance Asian ones) to enlarge the set of linguistic phenomena. Given the envisaged modular structure of MILE, a first phase of the encoding experiment has been dedicated to the creation of monolingual entries, while multilingual correspondences have been added in a second stage.

In this experiment, we wanted to simulate the scenario of independently built monolingual resources that are successively linked through multilingual transfer conditions. The following procedure has been adopted:

1. for each of the selected entries, we extracted the occurrences from various monolingual reference corpora (e.g. LE-PAROLE for Italian, BNC for English, etc.);
2. the extraction results have been organized in senses, with the help of existing monolingual dictionaries and computational lexicons (e.g. SIMPLE, WordNet, EuroWordNet/ItalWordNet, ComLex);
3. the relevant syntactic descriptions and the identified senses have been encoded according to the PAROLE-SIMPLE specifications (Lenci *et al.*, *2000*). The result has been a core of monolingual lexical entries described at the morphological, syntactic and semantic levels;
4. the various identified senses have been translated using bilingual dictionaries, and the translations have been revised by native language speakers;
5. on the basis of (4), the monolingual entries have been linked into bilingual entries, by focusing on the tests and actions that need to be expressed to establish proper multilingual correspondences.

Following the PAROLE-SIMPLE model, each monolingual entry has been described in terms of three interlinked entities, i.e. Morphological Unit (*MU*), Syntactic Unit (*SynU*) and Semantic Unit (*SemU*), which encode respectively the morphological, syntactic and semantic relevant information. In the SemUs, the various types of information available in the SIMPLE model (e.g. ontological types, examples, domain information, semantic features, semantic relations, thematic roles, selectional restrictions of the arguments, etc.) have been exploited to provide a formal characterization of the selected senses of the lexical entries. On the other hand, we focused on the necessary extensions and enrichment of the original model, especially in the prospective of the jump at the multilingual level.

A particularly critical issue both at the monolingual and multilingual level is represented by the dominant role of multiword expressions and collocations. These form a kind of lexicographic "no-man's land", which can not be easily captured with the expressive resources of standard computational lexicons. In many circumstances, it is also difficult to organize this highly context-dependent information within the main senses articulation of each word. The border between the purely lexical idiosyncrasy and the possibility of extracting useful generalization is a very thin line, whose effective characterization is nevertheless an important demand in multilingual computational lexicography.

An interesting and prototypical case is the Italian noun *colpo*, which is usually translatable with the English equivalent "blow" and "stroke". However, in many cases, a more specific translation is needed, depending on the surrounding linguistic context in which this noun appears. For example, when we find *colpo* in the common context: *Colpo+ di +*INSTRUMENT ([$_{NP}$[$_{N}$colpo]] [$_{PP}$[$_{P}$di] $_{NP}$[$_{N}$X]]]), we usually translate it with *stroke*, but:

if INSTRUMENT={frusta (*whip*)} then *colpo=lash*
if INSTRUMENT={falce (*sickle*)} then *colpo=sweep*
if INSTRUMENT={testa (*head*)} then *colpo=header*
if INSTRUMENT={tacco (*heel*)} then *colpo=heel print*
etc..

In the same way, when *colpo* is followed by an adjective it can be generally translated using *blow*, but:

colpo mancino = *an underhand blow*
colpo gobbo =*a stab in the back*
colpo basso = *a hit below the belt*

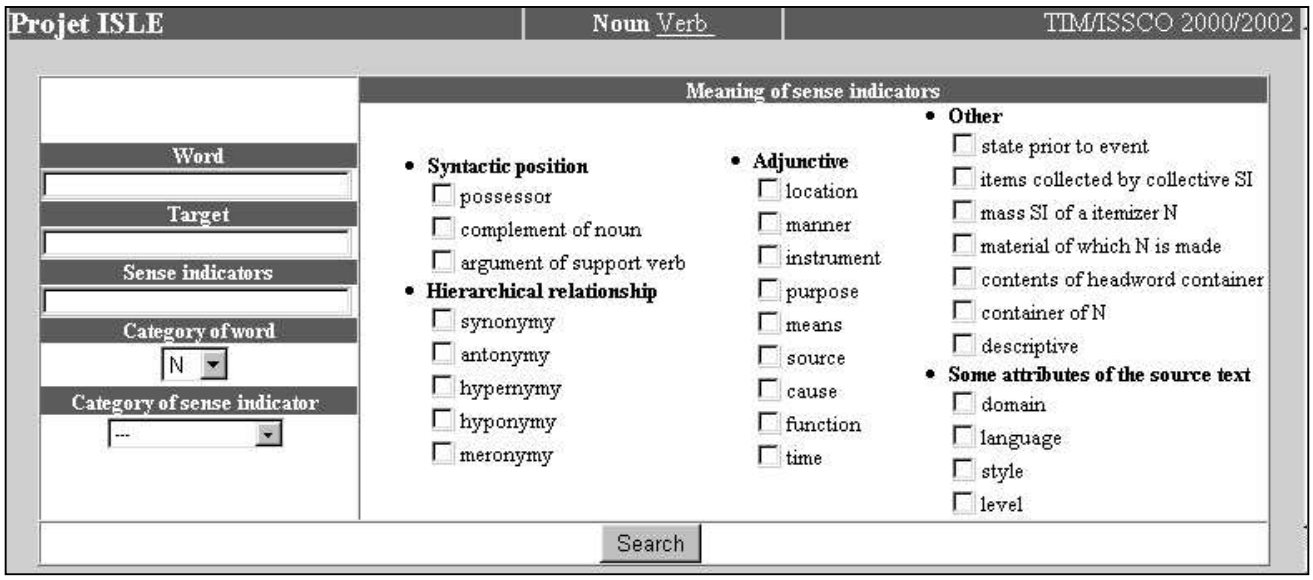

Figure 2: The *sense indicators* database

To deal with all these different and idiosyncratic situations, we had to create some specific SynUs at monolingual level:

synU: "colpo_di_frusta"
description: colpo+[prep="di"]+[N=lex] where [lex]="frusta"
……
synU: "colpo_di_tacco"
description: colpo+[prep="di"]+[N=lex] where [lex]="tacco"
……..
synU: "colpo_mancino"
description: colpo+[adj=lex] where [lex]="mancino"
etc...

Then, the *multi*-MILE layer we must simply record the correspondences between Italian and English SynUs:

Mult_Usyn: <colpo-di-frusta_lash>
Italian_Usyn: "colpo-di-frusta"
English_Usyn: "lash"

Mult_Usyn: <colpo-di-tacco_heel-print>
Italian_Usyn: "colpo-di-tacco"
English_Usyn: "heel_print"

Mult_Usyn: <colpo-mancino_underhand-blow>
Italian_Usyn: "colpo_mancino"
English_Usyn: "underhand_blow"

The correspondence established only at syntactic level is the most simple and direct, but we have to specify the whole range of transfer situations - between all the different layers of lexical description - for which we have to establish links. We are designing a model which provide us a set of explicit *lexical objects* (see section 4.2), that can be used to create new syntactic positions, new arguments, to constrain semantic and syntactic information via a powerful yet simple *lego*-mechanism of tests and actions.

4.1.2. A database of sense indicators

A second important task in the investigation and discovery of the lexical dimensions relevant for multilingual transfer conditions has been to create a database of lexicographic sense indicators. The goal was to identify and classify the 'clues' given by the lexicographer to the bilingual dictionary users in order to guide them to the most appropriate choice of equivalence in the foreign language. The database should therefore offer the answers to the following questions:

- what type of information is used in multilingual dictionaries to establish translation equivalents?
- how to classify this information in a way that reflects the lexicographically relevant facts (LRF)?

Answers to these two questions are of great help in leading to the formalization of the implicit information available in current bilingual dictionaries, and in formulating recommendations regarding effective transfer conditions in machine-assisted translation. Figure 2 shows a screenshot of the database Web-based GUI.

We extracted automatically from an English-French dictionary the source word with its syntactic category, the target words and the sense indicators (SI). We then classified the SIs with regard to 5 LRFs, namely:

1. whether the SI fills:
    a. a subcategorized position (subject, complement, etc.);
    b. an adjunctive/modification position (manner, instrument, etc.) vis-à-vis the headword;
2. whether the SI specify conditions on the headword, such as:
    a. semantic relationships (synonymy, hyperonymy, hyponymy, etc.);

b. textual information (style, level, language variety, etc.);
c. morphological information (number, gender, etc.);

For verbs of a bilingual dictionary, for example, statistics are as follows:

	Subcategorized pos.	Adjunct pos.	Hierachical rel.	Textual inf.
Intrans. verbs	355	44	656	166
Trans. verbs	1149	44	275	284

This figure tend to show that the two most relevant types of information concern the subcategorized positions and then the hierarchical relations. This can be compared to other categories and give interesting information about what is relevant from the point of view of the lexicographer.

Other interesting queries from the lexicographic point of view are:

- what properties do the verbs share which are encoded with the same type of LRF
- what types of verb are encoded with, for instance
  - SIs that have the pattern ' ' by N" expressing a manner
  - all SIs that contain a specific string like " US" ,' ' person' ' , etc.
  - all SIs that are adverbs, etc.

The database of sense indictors represents an important project internal resource, but it can possibly be developed into a more general supporting tool for lexicographers. The extracted sense indicators have been compared with the types of lexical information provided in PAROLE-SIMPLE, as well as with existing models of transfer conditions. This way, they have provided a useful help and guideline for the lexicographers/linguists in charge of the preparation of the sample multilingual entries, as described in section 4.1.1.

4.2. The formal data model for MILE

The ISLE CLWG is also working towards the design of the abstract data model for MILE entries. The main objective is to provide computational lexicon developers with a formal framework to encode MILE-conformant lexical entries. On the ground of the preliminary list of crucial information types present in monolingual lexicons and essential to establish proper multilingual transfer conditions, the ISLE group is setting up a lexicographic environment consisting of the following three main components (Figure 3 illustrates the overall framework):

1. an XML DTD formalizing MILE as an Entity Relationship model;
2. A first repositories of lexical data categories and shared objects, to be used to build in an easy and straightforward way MILE-conformant lexical entries.
3. the ISLE Lexicographic Station, which will map the MILE entity-relation model into a relational database, and will also include a GUI to input, browse and query the data in a user-friendly way.

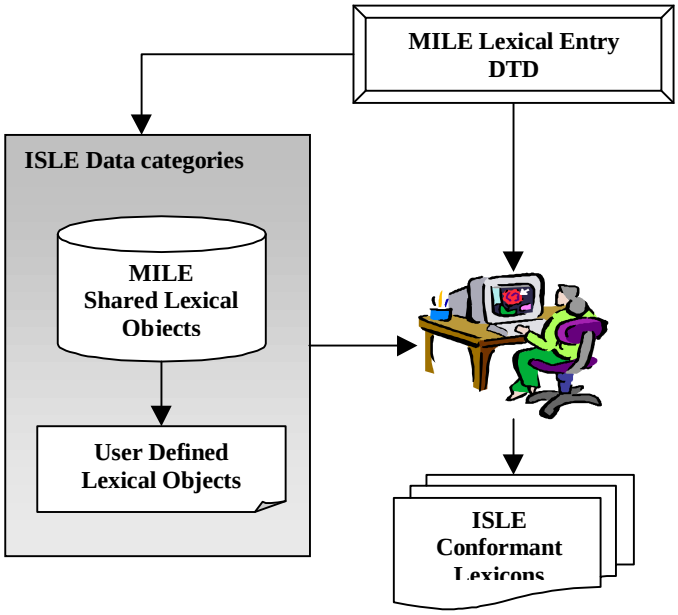

Figure 3: Formal Framework for MILE

On the formal side, a first low-level definition of MILE will be provided by an Entity Relationship model, encoded as an XML DTD. This will define the general constraints for the construction of multilingual entries, as well as the grammar to build the whole array of lexical elements (such as features, predicates, semantic relations, syntactic positions, etc.) needed for a given lexical description.

At a higher level, the ISLE recommendations will also consist of a first repository of *shared lexical objects*. These will include main syntactic constructions, basic operations and conditions to establish multilingual links, macro-semantic objects, such as lexical conceptual templates acting as general constraints for the encoding of semantic units.

For instance, at the multilingual level it is possible to identify a first set of basic operations that are at the basis of multilingual transfer tests and actions. This would include:

- adding to a monolingual lexical entry a new syntactic position (required for a given translation correspondence);
- adding to a monolingual semantic description a new semantic feature (required for a given translation correspondence);
- constraining the source-target correspondence to apply only if an existing syntactic position is realized by a certain type of phrase, etc.

This way, both an the monolingual and at the multilingual level (but with particular emphasis on the latter), ISLE intends to start up the incremental

definition of a more Objected-Oriented layer for lexical description. The defined lexical objects will be used by the lexicon (or applications) developers to build and target lexical data at a higher level of abstraction. Thus, they have to be seen as a step in the direction of simplifying and improving the usability of the MILE recommendations.

The ISLE Lexicographic Station is a development platform used to automatically generate a prototype tool starting from the MILE DTD. The aim of this prototype tool is to i) exemplify the MILE entry ii) make extensive use of already existing monolingual resources, and iii) eventually test the guidelines in a real scenario. This situation led us to define a lexicographic station development platform that guarantees the portability of the final prototype to the final specifications as well as to existing monolingual resources which will serve as the basic data for MILE (for a detailed description, cf. Villegas and Bel, 2002).

The lexicographic development platform has been designed as a tool generator which parses any DTD describing an Entity Relationship model in order to automatically (i) map the DTD into a relational dB and (ii) build up a user-friendly interface able to cover the most common lexicographic requirements –such as means to automatically load/download the database from/into external SGML/XML files.

Basically, the lexicographic station includes a generation module, a customisation module and a core web interface module which can be briefly described as follows.

The generation module automatically generates a relational dB out of a DTD. This benefits from the fact that a conceptual model expressed in terms of Entity-Relationship model can be easily mapped into a relational dB.

The customisation module allows the user to modify certain aspects of the dB at the time that overcomes some of the well known shortcomings of DTDs, such as typed references and type declaration.

The core web interface module consists of a series of scripts that allow to manage the dB with a friendly interface. Although user requirements differ from site to site according to in-house needs the tool comes equipped with a set of basic functionalities. A list of requirements includes:

- query and browsing facilities;
- import, export and migration of data;
- easy encoding of new data;
- test and validation of both the data and the model;
- customisation facilities;
- lexicographic tools such as type definition, class extraction and statistical facilities.

As in the case of the generation module, this web interface module acts on the model expressed in the DTD in order to make the necessary calculations to access, manipulate and display data from relevant tables.

5. Conclusions

In this paper we presented the preliminary results of the ongoing research activity of the ISLE Computational Lexicon Working Group. The design of a common and standardized framework for lexicon construction can lead to the optimization of the whole process of production of lexical resources: their creation, maintenance and extension, but also their reusability for different applications and tasks. It is critical to achieve the interoperability needed for effective integration, a precondition for a quallitative improvement in multilingual content processing technologies. A crucial aspect in establishing a real and broad consensus is played by communication and sharing of information among many groups active in the field. For this reason we involved also Asian collegues in the ISLE initiative, and we are exploring ways of establishing formal links with them.

6. References

Calzolari, N., Grishman, R., Palmer, M., Atkins, S., Bel, N., Bertagna, F., Bouillon, P., Dorr, B., Fellbaum, C., Gibbon, D., Habash, N., Lange, E., Lehmann, S., Lenci, A., McCormick, S., McNaught, J., Ogonowski, A., Pentheroudakis, J., Richardson, S., Thurmair, G., Vanderwende, L., Villegas, M., Vossen, P., Zampolli, A. (2001a). *Survey of Major Approaches Towards Bilingual/Multilingual Lexicons*, deliverable D2.1-D3.1, WP2-WP3, EAGLES/ISLE (IST-1999-10647).

Calzolari, N., Lenci, A., Zampolli, A., Bel, N., Villegas, M., Thurmair, G. (2001b). *The ISLE in the Ocean. Transatlantic Standards for Multilingual Lexicons (with an eye to Machine Translation).* In Proceedings of Machine Translation Summit VIII, Santiago de Compostela, Spain.

Calzolari, N., McNaught, J., Zampolli, A. (1996). *EAGLES Final Report: EAGLES Editors' Introduction*. EAG-EB-EI, Pisa, Italy.

Calzolari, N., Fillmore, C.J., Grishman, R., Ide, N., Lenci, A., MacLeod, C., Zampolli A. (2002). *Towards Best Practice for Multiword Expressions in Computational Lexicons*. In Proceedings of LREC 2002, Las Palmas, Canary Islands, Spain.

Lenci A., Bel N., Busa F., Calzolari N., Gola E., Monachini M., Ogonowsky A., Peters I., Peters W., Ruimy N., Villegas M., Zampolli A. (2000). SIMPLE: A General Framework for the Development of Multilingual Lexicons. *International Journal of Lexicography*, XIII (4), 249--263.

Lieske, C.,McCormick, S., Thurmair, G. (2001). *The Open Lexicon Interchange Format (OLIF) Comes of Age.* In Proceedings of the Eighth Machine Translation Summit, Santiago de Compostela, Spain.

Ruimy, N., Corazzari, O., Gola, E., Spanu, A., Calzolai, N., Zampolli, A. (1998). *The European LE-PAROLE Project: the Italian Syntactic Lexicon.* In Proceedings of the First International Conference on Language Resources and Evaluation (pp. 241--248), Granada, Spain.

Villegas, M., Bel, N. (2002). From DTDs to relational dBs. An automatic generation of a lexicographical station out off ISLE guidelines. In Proceedings of LREC 2002, Las Palmas, Canary Islands, Spain.