

A Library Information System (LIS) Based on UNL Knowledge Infrastructure

Sameh Alansary^{†‡}
Sameh.alansary@bibalex.org

Magdy Nagi^{††‡}
magdy.nagi@bibalex.org

Noha Adly^{††‡}
noha.adly@bibalex.org

[‡] Bibliotheca Alexandrina, P.O. Box 138, 21526, El Shatby, Alexandria, Egypt.

[†] Department of Phonetics and Linguistics
Faculty of Arts
Alexandria University
El Shatby, Alexandria, Egypt.

^{††} Computer and System Engineering Dept.
Faculty of Engineering
Alexandria University,
Egypt.

ABSTRACT

In this paper we introduce a prototype of Library Information Systems that uses the Universal Networking Language (UNL) as a means for translating the metadata of books. This prototype is capable of handling the bibliographic information of 1000 books selected from the catalogs of Bibliotheca Alexandrina (B.A.).

The paper sheds light, firstly, on the idea of sharing bibliographic information across languages; secondly, on the linguistic and computational challenges faced when trying to execute such an idea using UNL interlingua, and thirdly on the implementation of this innovative system.

Keywords

Universal Networking Language (UNL), Library Information System (LIS), Arabic UNL system, Universal Digital Libraries, UNL applications.

1. Introduction

Knowledge is for all, but to be indeed for all, it should be accessible for all those who seek it regardless of their mother tongue. Consequently, libraries as the organizers and heralds of this knowledge, adding value to it by cataloguing and classifying, should, in turn, be Universal; i.e. provide equality of access for all.

Today, Information Technology has converted the world into a global village and libraries, as part of this age, should make use of these technological advancements in achieving the Universality goal and quenching the generation's thirst for knowledge. This means that traditional libraries should change into well-equipped interconnected digital libraries.

But digital libraries, however, are language-dependent; in other words, their digital material will only be read in one's native language. Language dependency is, therefore, an obstacle facing the dissemination of knowledge, a fact that makes the knowledge in libraries, to some extent, limited [1]. Here comes the role of UNL in freeing knowledge of its linguistic dependency.

This paper first discusses in Section 2 how UNL as a universal language can be used for knowledge representation. Section 3 presents the ISAUC; the language center responsible for Arabic EnConversion and DeConversion. Section 4 introduces the UNL-LIS as a language-independent Library Information System and, finally, Section 5 concludes the paper and explores future work.

2. UNL as a Universal System for Knowledge Representation

UNL is as a artificial language used for representing the meaning of Natural Language (NL) sentences [10]. UNL is not limited to any particular domain and represents any content in the form of networks of interrelated nodes and arcs in which nodes represent the concepts present in the sentence, called the Universal Words (UWs), these can be further annotated with attributes to provide more information about how a concept is being used in a particular sentence, while, on the other hand, arcs represent the semantic relation between each pair of these concepts (UWs) [11].

Since it does not deal with NL input not as a block but rather analyzes it into concepts and relations, UNL can make searching within library catalogs semantic rather than orthographic. Moreover, the relations between concepts can help the automatic extraction of key words. Ultimately, UNL as an artificial language can help achieve the goal of Universality in Library Information Systems.

3. The Arabic UNL Language Center: a Story of Success

In July 2004, partnership with the Universal Networking Digital Language (UNDL) foundation have been established and an agreement has been signed in favor of Bibliotheca Alexandrina to host Ibrahim Shihata Arabic UNL Center (ISAUC), the center responsible for designing and implementing the Arabic component of UNL. The ISAUC has succeeded in building the infrastructure of the Arabic UNL system thanks to a team of highly qualified software engineers and computational linguists formed with the purpose of building the Arabic language tools and setting up the Arabic language server. A new team of linguists has also been trained with the aim of expanding the ISAUC.

The following tools have been built:

a) The UNL-Arabic Dictionary: Much attention has been given to the dictionary in order to make it suitable and in the required format to support the morphological, syntactic and semantic analysis and generation needed for both the Arabic EnConversion and DeConversion. In addition to the General Dictionary, two specialized dictionaries have been built one for the Encyclopedia of Life Support Systems (EOLSS) and the other for the Library Information System (LIS).

b) Arabic EnConversion Rules: Two versions of grammar have been developed for EnConverting Arabic texts automatically into UNL; General and Specialized. The General version can EnConvert any Arabic text, while the Specialized version is devoted to EnConverting book titles under the UNL-LIS project.

c) Arabic DeConversion Rules: These are the grammar rules responsible for generating Arabic sentences out of UNL

semantic networks (expressions). The current version of the grammar have been utilized in the DeConversion of over 13,000 sentences representing 25 documents from the Encyclopedia of Life Support Systems (EOLSS) and the results were satisfactory.

d) The Arabic Corpus: The Arabic UNL center initiated a project to build the International Corpus of Arabic (ICA); a realistic and substantial attempt to build a representative corpus of the Arabic language as it is used all over the Arab world. It depends on samples of written Modern Standard Arabic (MSA) selected and collected from a wide range of sources. The goal of the project is to collect and analyze 100 million words. Over 60 million words have already been collected, and the compilation is still in progress. The analysis phase is also in progress, a 200,000-word training corpus has been built including the morphological analysis, gender, number, word class, case, and root of each word. Such a project should support and advance research on the Arabic language, and help researchers explore Arabic texts more deeply. The ICA will also help update the Arabic-UNL Dictionary, the EnConversion grammar and the DeConversion grammar.

e) UNL Supporting Tools: A number of tools have also been built to help improve the various components of the UNL system. One of these tools is the UNL Integrated Development Environment (UNL IDE); a tool that enables users and developers to view the UNL semantic network, search UNL documents, write rules, check their syntax, and debug to watch the DeConversion and EnConversion output for the given rules and dictionary. A wide range of improvements is also under progress to enhance the UNL IDE and give it the power to deal with the forthcoming challenges facing UNL.

4. Building a UNL Library Information System (UNL-LIS)

The UNL-LIS is an application that allows for the retrieval of the metadata of books available in a library data store, it allows users to access this information in their own native language, regardless of the original language it is written in.

4.1. The Selection of Sample Book Titles

The most sophisticated information in book records are titles; hence, in order to UNLize the LIS, a sample of book titles have been selected from the B.A. catalog to be subsequently EnConverted. This sample included about 800 Arabic titles, 100 English titles and 100 French titles, selected according to certain parameters to represent the various phenomena that manifest themselves in books and their metadata. Some of these parameters are linguistic:

- Titles should include statements and questions.
- Titles should include verbal and nominal sentences.
- Titles should include short and long sentences.
- Some titles include subtitles.

Other parameters that would help test the search engine of the LIS are:

- Authors should be of varying degrees of fame.
- Some authors should have written more than one book.
- Some subjects should be covered in more than one book.

Table 1 shows statistical information about the number of words in the selected Arabic titles. The minimum number of words is 1 word while the maximum is 9. It is clear from the table that the great majority of the sample titles consists of 2 words; that was a deliberate choice to help the EnConversion process perform better. Nevertheless, some longer titles have been selected to examine the types of complexities that would show up in such titles.

1 word	2 words	3 words	4 words	5 words	6 words	7 words	8 words	9 words
51	333	183	80	67	37	17	10	23

Table 1: statistical information about the number of words in the sample titles

Chart 1, on the other hand, shows the various syntactic structures of the sample titles; three structures have undergone analysis: full sentences, compound phrases (such as a title and its subtitle) and single phrases.

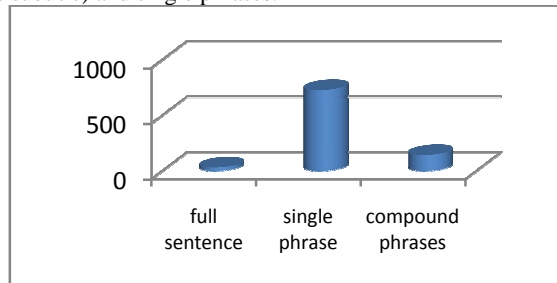


Chart 1: statistical information about the syntactic structure of titles

It is possible to think that dealing with book titles in the field of Natural Language Processing (NLP) is an easy task since they are, generally, short sentences or phrases; however, it is not. This fact will be demonstrated in the following subsection (4.3) discussing some of the linguistic and computational issues that stand in the way of EnConverting book titles.

4.2. The Workflow of the UNL-LIS

The workflow of the UNL-LIS starts by extracting the metadata of books from the MARC21 records manually and then verifying them. The metadata then undergo semantic analysis (EnConversion) to identify the UNL relations between the titles' words. The UNL expressions of these titles are then generated with the help of the EnConversion grammar and the Source Language-UW dictionary stored in the language server of each language. After this stage, a UNL specialist checks these UNL expressions for wrong relations or undefined UWs; if he/she finds a relation to be invalid, he/she would fix it by modifying the EnConversion rules, but if the problem had to do with an undefined UW, it would be defined by a UNL UW specialist and inserted in the dictionary. This process continues until the EnConverter outputs a valid UNL expression (see Fig.1). Afterwards, the output UNL expression is stored in the catalogs as a book record ready to be DeConverted into any Natural Language supported by UNL.

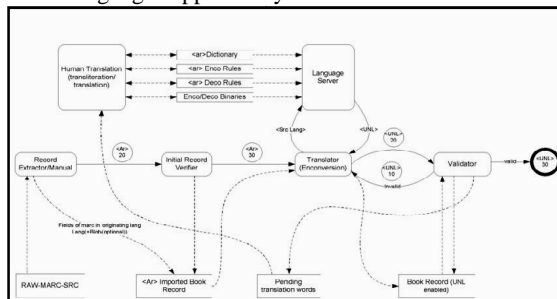


Fig.1: NL-UNL conversion (EnConversion)

After UNL expressions are verified and sent to the corresponding language center, they enter the Universal Words checker to check whether any of the incoming UNL expressions contains a new undefined UW, if yes, it would be defined in the UW dictionary of the target language found on its language server. The UNL expressions are then ready to be DeConverted by means of the UNL DeConverter and with the help of the Target Language-UW Dictionary and the

DeConversion rules stored in the corresponding language server. After the DeConversion into the target language is complete, a librarian acting as a NL-Validator checks whether the DeConverted title is suitable for the book. If the title is indeed suitable, it will be stored as a book record in the target language; if not, the librarian returns it to the UNL specialist to DeConvert it again until they reach a valid natural language book record in the desired target language (see Fig.2).

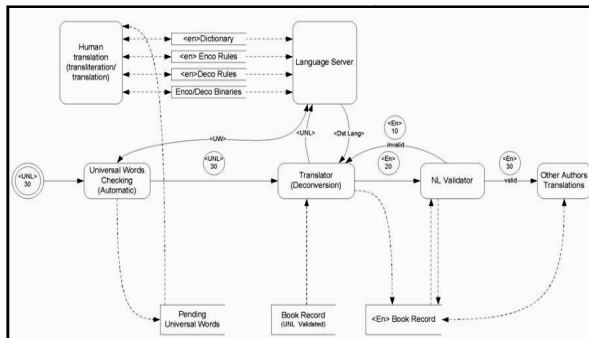


Fig.2: UNL-NL conversion (Deconversion)

4.3. Book Titles: Linguistic Issues

Book titles may be compared with other varieties of syntactically reduced language, such as headlines, chapter headings, all of which may defy traditional syntactic analysis [3]. And since book titles have not been studied before and since they share so many linguistic features with headlines, their linguistic features will be discussed as thus in this paper.

There is a long history of linguistic interest in the syntax of headlines being reduced compared to the norms of written language [2]; Syntactically, a headline “falls outside language proper” [4]. It has, however, long been realized that headline syntax is rule-governed in the same manner as other reduced forms of communication. Ellipsis, for example, is a common compacting technique used in headlines. Jenkins draws attention to other techniques including: “stacked nouns”, preference for short words, abbreviations and initials [5] among the many linguistic features that characterize book titles.

Moreover, book titles share certain pragmatic functions with headlines in providing an initial introduction to attract the reader’s attention or communicate certain information. Thus, authors try to write titles that are ambiguous or confusing to arouse the reader’s curiosity and encourage him/her to read the whole content of the book.

4.3.1. Arabic-UNL Dictionary: Issues on the Choice of Universal Words

A specialized version of the Dictionary and Grammar have been built in order to EnConvert the selected book titles. The LIS Specialized Dictionary has been built by, first, extracting the words in the selected titles and choosing the UWs that accurately represent their meanings. Then, assigning them the appropriate linguistic features. Nevertheless, some challenges have been encountered in the course of choosing the appropriate UWs some of which are:

- Some titles have already been human-translated as a whole to convey the meaning the translator saw fit the book. These titles were UNLized as usual but, naturally, the resulting translation did not exactly match the human-translation. Therefore, in order to help people already familiar with these usually renowned books and translations, the human translation has also been defined in the dictionary as a whole concept with the tail (iof>book). Some examples are “عذاب الحب الضائع”, “عيد ميلاد جديد”, and “ضجة فارغة” the original titles of which were defined as “Love’s Labour’s Lost (iof>book)”,

“A Different Kind of Christmas (iof>book)” and “Much Ado About Nothing (iof>book)” respectively.

- Other words were culture-specific; i.e., they have no English equivalent to act as the Head of the UW; hence, they had to be transliterated. Some examples are the words: “حديث” “the sayings of prophet Muhammad” and “حجاب” “a headscarf worn by Muslim women” both of which have no appropriate English equivalents and were, consequently, defined as “Hadith(icl>oral tradition)” and “Hijab(icl>veil)”.
- Some words were used Metaphorically; for example: “أشواك” “the thorns of peace”. In this example, it is difficult to determine the appropriate Head Word equivalent to the word “أشواك” and whether it should be translated literally into its metaphoric meaning “thorns” or into its intended meaning “disadvantages”.

The number of unique UWs in the LIS dictionary is about 1750 including those of titles, authors, publishers, and subjects. The overall size of LIS dictionary so far is 3000 entries representing 1300 UWs.

4.3.2. Enconversion: Formal and Computational Linguistic Issues

Book titles are a proper subset of block language; the language used in telegrams, headlines, notes, advertisement ... etc. Block language is made of the word or phrase rather than the clause or sentence [4] and the grammar utilized in this form of language differs from full grammar in a rather well defined way [9].

In the process of LIS UNLization, a specialized version of EnCo grammar have been developed to deal with the linguistic phenomena that present themselves in book titles.

4.3.2.1. Identifying the Semantic Relations Between the Components of Compound Titles

A common structure among the sample titles has been two noun phrases juxtaposed on either side of, most usually, a colon, a full stop or a dash.

As a rule, titles and subtitles are syntactically independent of each other but semantically interdependent [6]. However, the semantic relation between a title and its corresponding subtitle may not be always the same. Let us consider the book titles in (1) as an example.

- (1)
 - a) الليزر: بين النظرية والتطبيق
Laser: Between Theory and Application.
 - b) المعماريون العرب: حسن فتحي
Arab Architects: Hassan Fathy.
 - c) سكان الإسكندرية: دراسة ديموغرافية منهجية
Inhabitants: a Systematic and Demographic Study.

(1a) is the simplest case where the colon does not add any meaning; i.e., if the colon is removed, we will end up with the same meaning without any distortion or omission. The syntactic structure of this title as constructed of two parts (as shown in (2)) implies that the second phrase is a prepositional phrase (PP) complement of the first phrase; Noun phrase (NP) which explains why the colon does not add any further meaning. This phenomenon is described formally in (3):

- (2) NP → DET-N
PP → P - NP[NP-CO-NP]
- (3) >{N,&@def,fix}{PP,fix:rel:aoj:}P2;
>{PP,fix}{N,fix:rel:obj:}P2;

As the UW (between(aoj>thing,obj>thing)) of the Arabic word “بين” will be the main entry of this title, the relation between it (the main entry of the sentence) and the UW (laser(icl>beam)) of “الليزر” is an aoj relation, and the coordinated clause “النظرية

”والتطبيق“ is an obj relation. This is the correct analysis when the word class of the word directly after the colon is a preposition. Let us now apply the same approach on (1b). If the colon were to be removed from the title, the title would be “المعماريون العرب الحسن فتحى”, “Arab Architects Hassan Fathy” which does not sound like a natural Arabic sentence, and is difficult to understand, if possible at all. However, if the title is read with the colon, it clearly conveys the meaning “حسن فتحى كأحد المعماريون العرب”, “Hassan Fathy as one of the Arab Architects”. Although word order (syntactic structure) in (1a) and (1b) is more or less similar, their semantic meaning is different because the colon in the second title introduces a proper name; [NP[N[حسناً فتحى]]] as in (4):

(4) NP → NP[DET-N_{plural}]- ADJP[ADJ]
NP → NP[N_{proper name}]

Therefore, if a title and its subtitle follow the syntactic structure in (4), they should be analyzed as (N_{proper name} is one of [NP[DET-N_{plural}]]), and linked by an *iof* relation. This is expressed formally in (5):

(5) <{N,HUM,fix,&@pl,&@colon,rel::iof:}{N,fix,PROPN,SIN}P10;

Let us consider the example in (1c), we will notice that if we remove the colon, the title will be: “سكان الإسكندرية دراسة” ‘Alexandria’s inhabitants: a systematic and demographic study’ which is clearly an ill-formed sentence.

What is the role of colon in this title, then? To understand this title, it might be the case that the Arabic native speakers unconsciously change the title into “دراسة ديموغرافية منهجية لسكان الإسكندرية” ‘a systematic and demographic study of Alexandria’s inhabitants’, i.e. two processes have been made :

- Reversing the two phrases.
- Inserting the preposition “ل” before the second phrase (after the reversal).

Thus, the EnCo rules try to formalize these two processes by, firstly, analyzing each part syntactically in an independent manner. Then, if the two phrases were found to follow the syntactic structure shown in (6), the EnCo grammar would push the head of the second phrase to assign a “mod” relation to the head of the first phrase, by that it perceives the two phrases as reversed and supposes that the “ل” is inserted between them, this is shown in the formal rule (7).

(6) NP → N-NP
NP → N-ADJP[ADJ1-ADJ2]

(7) <{N,&@colon,<mod::mod:}{N,<aoj:::}{^ADJ}P10;

4.3.2.2. Elliptic Constructions

An elliptic construction is a linguistic construction that lacks an element that is, nevertheless, recoverable or inferable from the context; i.e., it is a sequence of words in which some words have been omitted, but because of the logic or pattern of the entire sentence, it is easy to infer what the missing words are. For example, *Fire when ready*. In this sentence, “you are” is implicitly understood, as in “*Fire when you are ready*.”

Titles display a high frequency of ellipsis and often omit a number of words and/or use words in a shortened form [9]. Our analyzed sample of titles contains two types of elliptic constructions; namely, elliptic phrases and elliptic coordination

Elliptic Phrases:

Nominal sentences in Arabic consist of two parts; Topic and Comment. The title in (8), begins with the preposition “حول” “about”, thus, (8) is not a sentence but rather a phrase. In this title, the prepositional phrase “حول الوحدة الثقافية العربية” acts as the predicate for an omitted subject referring to the book itself.

(8) “About the Arab Cultural Unity” حول الوحدة الثقافية العربية

Semantically, the title must have a Topic. Hence, UNL first expresses this omitted Topic using the “null” UW by the EnCo rule in (9a). Then, An *aoj* relation relates this “null” UW to the preposition “حول” “about” by the rule in (9b).

(9) a- : (SHEAD) "[Null]:#INSE,mor" {PREP,^#INSE,mor} (N)P2;
b- >{Null,fix,^COO::aoj}{PREP,fix,^>aoj:rel}P2;

• Elliptic coordination:

Elliptic coordination represents a major problem since the EnCo grammar cannot handle it automatically. Consider the title in (10).

(10) أسس الجغرافية العامة: الطبيعية والبشرية
“The Fundamentals of General Geography: Physical and Human”

On first consideration, one would suppose that the components of the coordinated phrase “الطبيعية والبشرية” “physical and human” qualify the word “أسس” “fundamentals”; but in fact they do not, they rather qualify the word “الجغرافية” “geography”. The science “General Geography” includes many branches two of which are “Human Geography” and “Physical Geography”. Hence, the title here states that the book will discuss the fundamentals of only these two branches of “General Geography”. In this case, the coordinated clause will not include two adjectival concepts but rather two nominal ones; namely, “Human Geography” and “Physical Geography” these should then be connected semantically to “General Geography” through an *icl* relation. This analysis, however, cannot be reasoned out automatically and, therefore, the input title has to be manually modified into “أسس الجغرافية العامة: الجغرافيا الطبيعية والجغرافيا البشرية” “The Fundamentals of General Geography: Physical Geography and Human Geography” in order to force the EnConverter into selecting the nominal concepts from the dictionary rather than the adjectival ones, this selection, in turn, facilitates the designation of the *icl* relation.

4.3.2.3. Metaphors

A metaphor expresses the unfamiliar in terms of the familiar; a figure of speech in which a word denoting one subject is used in place of another to suggest similarities between the two [8]. Metaphorical expressions represent 7% of the sample book titles. They constitute a problem because it is difficult to determine the concepts referred to by the words in a metaphoric title and the semantic relations between them.

The problem of choosing the suitable UW equivalent to a word used metaphorically is demonstrated in the example (11a). In this example, the appropriate sense of the word “زهرة” “flower” is not easily determined in (11a); the word “زهرة” out of this metaphorical context refers to the familiar concept of “flower” but in (11a) such a concept would be inappropriate. This is because the word “عمر” “life” indicates that the meaning of “زهرة” in this particular sentence is “prime”; the period or phase of ideal or peak condition.

(11) a- زهرة العمر. "The prime of life".
b- البحر يحاكم سمكة. "The sea sues a fish".

This situation can be handled formally by using an attribute that marks these two words in the dictionary as two constituents of a metaphor; P1_METAPHOR and P2_METAPHOR, as in (12):

(12)?L{N,P1_METAPHOR,^&@def,mor}{N,P2_METAPHOR,&@def}P2;

The P1_METAPHOR attribute indicates that the word “زهرة” is the first part of a metaphorical expression, while P2_METAPHOR refers to the second part “العمر”. The attribute P1_METAPHOR is assigned to both senses of the HeadWord “زهرة”; “flower” and “prime”, with higher priority given to the more common one; “flower”, to force the EnConverter into

selecting it in most cases. On the other hand, if this P1_METAPHOR "زهرة" is followed by a P2_METAPHOR "العمر" it should change to suit the metaphoric meaning of the title.

The second problem has to do with choosing the appropriate semantic relations that links the words of a metaphoric title. This problem is represented in the Selectional Restrictions of verbs; these are the restrictions a verb puts on its arguments [7] and they are crucial to determining the relations between the verb and the other words in the title. In (11b) the verb "حاكم" has two arguments; a subject (syntactic agent) and an object. The selectional restriction that the verb "حاكم" puts on its arguments is that both its subject and its object should be human beings as in (12).

(12) V.[NP_{+human} - NP_{+human}]
 Subj - obj

However, in (11b) neither the agt "البحر" "the sea" nor the object "سمكة" "a fish" of the verb "حاكم" "sue" are Humans, but the author of the book chose to use them as a metaphorical personification to attract the reader's attention. Hence, the verb's arguments should be modified in the Grammar to include Humans and creatures that can assume human roles. The EnCo grammar expresses the modified verb arguments by the rules in (13) where (13a) connects the agt "البحر" "the sea" with the verb "حاكم" and (13b) relates the obj "سمكة" "a fish" with the verb.

(13) a. >{N,AN,HUM_ROL,fix::agt:}{V,INT,fix:rel}P2;
 b. >{N,AN,HUM_ROL,<agt,fix::obj:}{V,INT,fix}P2;

4.4. The implemented UNL- LIS

ISAUC has completed designing and implementing a prototype LIS capable of translating the metadata of books into the six official United Nations languages (Arabic, Chinese, English, French, Russian, and Spanish) in addition to Portuguese. As shown in Fig.3, the interface of the system is divided into three main sections; a Language Bar (at the top) that displays the available target languages to the user to choose from, a Side Panel containing the main options in the application (Browse, Edit, and Translate), and the Content Panel.



Fig.3: The opening page of the LIS

This system enables users to either browse the available book titles as shown in Fig.4, or search within the database using the book's information: its publisher, author, classification... etc, in addition to keywords to limit the possible results (Fig.5). Upon choosing any of the retrieved results, the full metadata of the book is displayed.



Fig.4: Browsing available titles



Fig.5: Searching within the database

Moreover, the application can show statistical information on the stored books and their metadata such as the language they are written in, as shown in Fig.6.

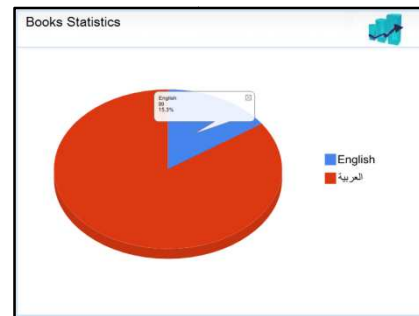


Fig.6: Statistics about the language books are written in

An editing option has also been designed for librarians to facilitate the cataloging of books. Using this option the librarian can add an authority, merge existing authorities, add or edit books metadata and link books with their human translations, if found. Fig.7 shows how librarians can create a new authority value, add a new synonym value, and change the master authority value.

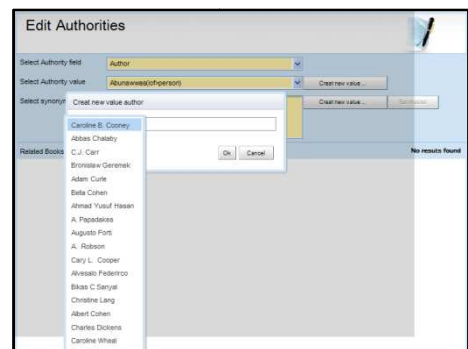


Fig.7: Adding authority values to the system.

Fig.8 shows how books' metadata are added or edited.

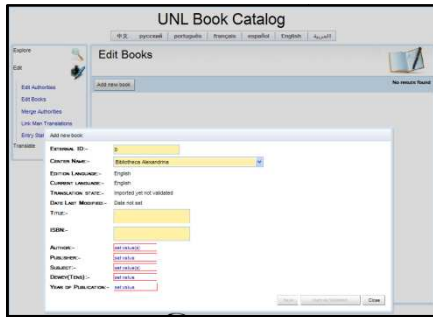


Fig.8: Editing books metadata

The application can also merge synonymous authority values that are different due to the different conventions of writing, abbreviation and name ordering as shown in Fig.9.

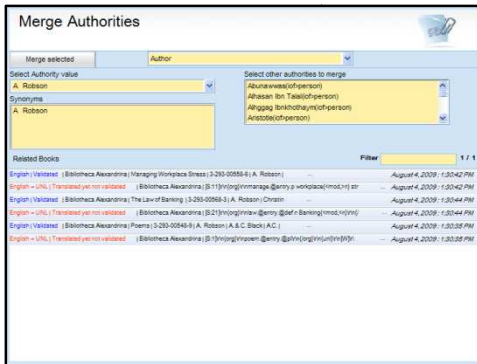


Fig.9: Merging synonymous authorities

The system can also provide the user with statistical information about the number of books that have been stored, translated and verified, and statistical information about their authors, publishers, subjects...etc. Fig.10 shows an example of these statistics.

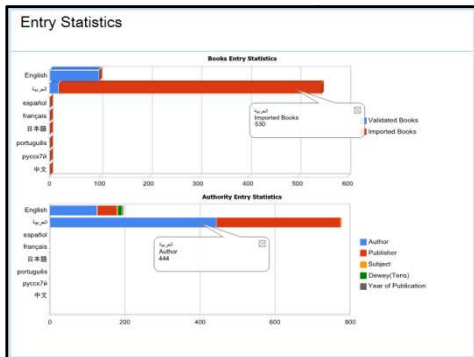


Fig.10: Entry Statistics

The Translation Process:

As explained in section 4.2, the translation process takes place over two steps. The first step is concerned with EnConverting the source language into UNL (the UNLization process) while the second is concerned with DeConverting book entries from UNL into the target natural language. Fig.11 shows the section devoted to the translation of the bibliographic metadata of books. The two steps have been implemented as follows:

a- **The EnConversion Process:** After librarians finish the data entry process, the EnConversion into UNL expressions (UNLization process) begins, and the resulting expressions are validated on the hands of linguists.

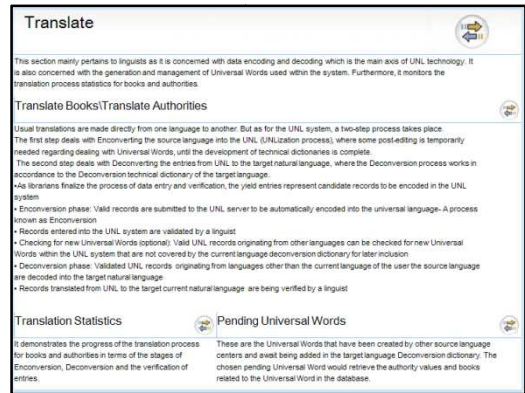


Fig.11: The Translation process front page

After validation, the UNL expressions are sent to the target language center which, in turn, checks them for undefined (Pending) UWs. These are Universal Words that have been defined by other language centers and are waiting for inclusion in the target language DeConversion dictionary. Fig.12 shows how the system handles these Pending UWs.

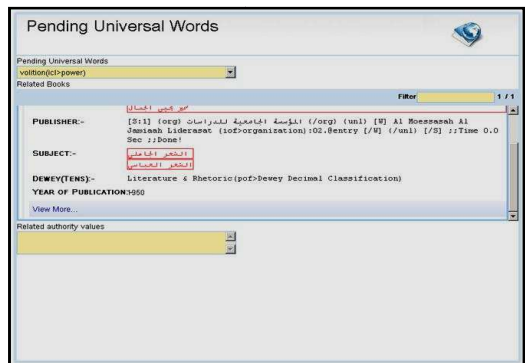


Fig.12. Pending Universal Words

b- **The Deconversion Process:** After adding the Pending UWs to the target language DeConversion dictionary. The DeConversion process begins. Afterwards, the output natural language is verified by a linguist.

Finally, the system provides users with statistics about the EnConversion and DeConversion processes showing the translation status; such as the percentage of data not yet EnConverted, EnConverted but not yet validated, and the DeConverted data . . . etc. Fig.13 shows an example of these statistics.

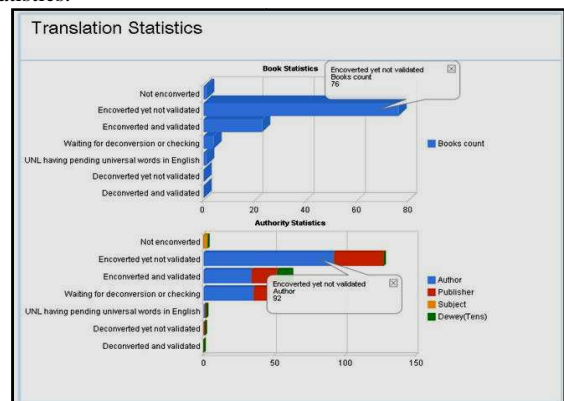


Fig.13: Translation Statistics

5. Conclusion

In an ideal world, users should be able to access library collections of any particular country in their own native language. This paper presents an innovative interlingua approach to achieving this goal of knowledge dissemination over the whole world; this approach is UNL. UNL is a promising interlingua that allows for the representation of Human knowledge through a language-independent means. It accurately expresses any specific concept in any particular language in a form that is comprehensible for computers and humans of different linguistic backgrounds alike and without any ambiguity.

In this paper, we have presented the contribution of the ISAUC in the design and implementation of an LIS application based on UNL. The UNL-LIS application allows users to search within UNL-based library catalogs and receive the results for their search both in their own native language. Book titles are not easily UNLized as some first thought; they display a high degree of ambiguity in the form of reduction, punctuation, prepositions and word sense...etc. These ambiguities are dealt with in a specialized version of the EnCo grammar. However, many more book titles have to be tested on this version of EnCo grammar, in order to discover any new undealt with linguistic phenomena that manifest themselves in book titles, before the UNL-LIS can be considered a powerful language-independent Library Information System.

6. References

- [1] S. Alansary, M. Nagi, N. Adly, "Towards a Language-Independent Universal Digital Library", *The Second International Conference on Universal Digital Libraries (ICUDL 2006)*. Alexandria, Egypt 17-19 November 2006.
- [2] P. Lennon, "Allusions in the Press: An Applied Linguistic Study". Berlin. 2004.
- [3] H. Straumann,. "Newspaper Headlines, A Study in Linguistic Method". London 1935.
- [4] O. Jespersen, "The philosophy of grammar". London: Allen & Unwin, P311. 1924.
- [5] H. Jenkins, "Train Sex Man Fined: Headlines and Cataphoric Ellipsis". 1987.
- [6] F. Brunot, "La pensée et la langue". Paris, P7. 1953.
- [9] J.J.Katz, "Semi-Sentences" in the structure of language P411. (Englewood Cliffs,N.J. :1964)
- [7] A. Radford, "Transformational Syntax: a student's Guide to Chomsky's Extended Standard Theory", Cambridge University Press. 1981.
- [8] S. Chung, "Source Domains as Concept Domains in Metaphorical Expressions". *Computational Linguistics and Chinese Language Processing*, Volume 10. 2005.
- [10] H. Uchida, "Knowledge Description Language", *Semantic computing workshop*, Tokyo, Japan, 2003.
- [11] H. Uchida, M. Zhu, "UNL2005 for Providing Knowledge Infrastructure", *SeC2005 Workshop*, Chiba, Japan, 2005.