ISO/IEC JTC1 SC36
Information technology for Learning, Education, and training

Title:

SC36 Cartago Alliance

Source:

AUF Liaison with WG1

Project:

NA

Document type:

Working document

Status:

FYI

Date:

2006-08-20

Action ID:

FYI

Distribution:

P, O, & L Members, WG Conveners

Cartago alliance

Presented by: Henri Hudrisier, Mokhtar Ben Henda, Elisabeth Blanchon
Contributions & adaptations: Mokhtar Ben Henda, André Le Meur, Michel
Porchet.

A) Introduction: the establishment of an international Consortium

The AUF terminology workshop at WSIS, has decided to found an association to work at the creation of terminological resources for multi-lingual and multi-cultural aspects in the based standards established especially by ISO/TC37: Terminology and linguistic resources.

A steering committee was appointed with responsibility to determine the name of the association, the structure of the association, and the scope and mission of the association. The steering committee will report to those involved before the end of December 2005. The provisional name of the project is Cartago and the website will be at: fabulameda@mshparisnord.org (adress for exclusive use of the project).

The steering committee is formed with individual participants instead of institutional representations:

Jean-Michel Borde	President@ailf.ass.fr
Elie Rafidinarivo	erafidin@refer.mg
Henri Hudrisier	hhudrisier@mshparisnord.org
Jin Gon Shon	jgshon@knou.ac.kr.
Rachid Zghibi	rachid_zghibi@yahoo;fr
Liddy Neiru	liddy@sunriseresearch.org
Daniel Prado	d.prado@unilat.org

Michel Porchet (mporchet@mshparisnord.org) and Joseph Tientcheu (tientcheuh@aol.com) take care of the moderation of the realm of TEI and Cartago.

We will discuss the following proposals:

1) The structure of the project and the regulation of the alliance:

- Structure: the institutions implied are: AUF (by the mean of Henri Hudrisier, liaison « AUF-SC36 »); MSH Paris Nord (Program 3 of the theme 4 of the axis 1 (MP, HH). This list will be completed regarding eventual adhesions (inscriptions).
- Regulation and respective role, we have to define soon a functional way and a minimum of organism (presidency, desk, technical tasks, contents, counsel.

2) Project scope:

The AUF terminology workshop at WSIS identified the need for support for multi-lingual and multi-cultural diversity in terminology to be used in e-learning standards. It was decided to form an association to work towards appropriate specifications, based

on established standards. The Cartago project wants to be complementary to works already started by other institutions implied in standardization of e-learning.

The partners of the Consortium have for one obligation to define together a model of application of TMF and then to self appropriate locally the conditions of contribution and use in order to respect the future Chart of Cartago.

The Cartago project intend to precise his purpose to serve a interoperability fully multilingual, multicultural, open on a prospect of the evolution of e-learning, and not to limit itself to consensus of experts of the different fields in standardization.

The target of the Consortium is to start with concrete questions like:

- To situate a concept in a system (network) of concepts.
- To verify the pertinence of the ontology as (implement tool) structuring tool of the elaboration of terminology in e-learning.
- Create a consensus and a network of critic mass of terminological data.
- To self organize to manage and share (exchange) contributions of all, examining reel needs.
- To refer to standards and, without a series of automated tests, to assure the collaboration of experts enable to guaranty the pertinence and conformity of the works.
- To do with the Consortium a reference of the multilingual and multi cultural terminology.

The Cartago Project will define in a Chart his own ethic.

3) Decisions for the method development:

The Terminological Mark-up Framework/TMF(ISO16642) is a structural scheme allowing the entire categories of standardized terminological data (ISO12620) conforming to the methods of ISO704. Following a common principle to all modern systems of structured information, TMF does not foresee any particular technological configuration. XML Schema is open to a maximum of possible applications assuring and defining interoperability. The organized data in TMF (which also can be written in TML or GMT) can be considered in technologically much diversified configurations (interfaces in different languages, application in centralized DB, sharewares contributing to terminologies, ontological constructions, and so on…).

This model has to be configured for the terminology of the e-learning. That's what is proposed by the models explained in the annex (complete model and models in lighter versions for the experts of e-learning who neither are linguists nor terminologists, experts have to complete these models).

The proposed categories of data have to be examined section by section to verify their compatibility with founding partner languages, and what eventually may be accumulated in provision for the languages of the founder zone (the Chinese, the Japanese, for example). Strong Incitation for a discussion with the Arabic and the *malgache*.

The data model is the key of such interoperability. That's why, despite its very technical character, it is presented in this document.

4) Decisions around the structural organisation for the future technical development

- Description of functions and tasks which could be realized in a quite reasonable way (discussion, vote…)

- Asking for offers and propositions in direction with the founding members and their environment, ideas for bi- or multicultural projects, eventually in association with the works of the Dublin Core, the W3C and the IMS.

5) Decisions about the structural organisation for the future contents development and their management

-Description of the content which we reasonably think to be able to realize (discussion, vote...)

-Asking for offers and propositions in direction with the founding members and their environment.

-Strict implication with the SC36 works, especially beyond the actual WG1, should help improving it to serve the needs of semantic terminals of other WGs.

-Define in common with certain experts of the SC36 implicated beyond the extents of the WG1 (WG2, WG3, WG4, WG5, WG7) the types of terminologies which could be useful to them and to start them with the idea to enlarge linguistically and culturally (for example: descriptions of the institutional environment, of certifications and of levels).

B) Some remarks of "the team Cartago project" for the WG1:

NB: We distinguish well following the presentation "the Cartago team project", which engages only partners of the MSH Northern Paris and those which collaborate directly for the development of the model project of "Cartago Alliance", founded in Tunis as reminded above.

1) The report Cartago/Termium: Globally, the implementation of Cartago does not at all discredit the interest towards Termium. However 3 notes are to be forwarded:

(a) Cartago is interested with terminologies for e-learning as currently developed. Therefore the role of Cartago is an upstream role compared to Termium which takes into account the terminologies of different JTC1 SCs after their standardization by SCs
(b) Cartago has ambition to work on a broad panel of languages and writings
(c) Cartago alliance defined certain numbers of ethical and economic rules, in particular we would like that the semantic and terminological reference frames of the e-learning can be developed for free and under free environments (thus a constraint

more ambitious than the only exemption from payment for "manufacturing standards". (These same "manufacturing standards" note indeed today that the trade of the contents and of the tools of the e-learning and E-training would be penalized if each recourse to a terminological reference of the field were to be paying.)

2) Vision and proposal for liaison with RG1/WG1: *(notes by Henri Hudrisier, in charge of liaison with WG1/RG1).*

We must note that the terminology which this liaison needs requires a description of the various commercial, industrial and administrative authorities of different NBs and LOs. This description requires obligatorily a taking into account of linguistic and cultural diversities. In addition, as noticed and initiated in Turku, this liaison requires the definition of a certain very general number of concepts on the field who exceed the implicit ones shared by the community of the experts of the SC36.

C) Terminology, largely multilingual: the proposal of "the Cartago team project"

Today the achieved progress is primarily technical: the project team consolidated the first model and especially improved and consolidated a model of structure of data.

While being allied with the University of Rennes (Andre LeMeur) and with that of Paris Jussieu (Elisabeth Blanchon), both experts within TC37, we largely sophisticated and developed the first model presented in Tunis

1) Fundamental assumption of the multicultural and multilingual management of Cartago: it is based on the pragmatic observation that we all control only basic English and that we have the practice of a collegial control at least bilingual "en & fr".

    a. According to current decisions of SC36 WG1, we have a model in which the conceptual level is defined in "en & fr".
    b. As in the language section, all the languages are equal. Thus all the languages have the possibility of giving their definition including "en & fr", have the possibility of giving to this level their cultural subtleties.
    c. If a linguistic culture wants to propose in Cartago a conceptual characteristic, a concept of which it thinks that it should either be clearly identified to allow interoperability or to be a concept which can then have vocation to be adopted by other linguistic cultures (or regional) this concept must be proposed on the level of the concept and be defined in "en & fr". Moreover, in the linguistic section, this concept must be of course defined in the source language.
    d. In Cartago, a definition in a given language will be always posted and accompanied by the definition in "en & fr"

2) What does it mean to be semantically and linguistically inter-operable in e-learning?

    a. This helps a basic user (relative of pupil, student, professional orientator) to get information in its language about for example a specific training just as it is in various cultural and linguistic contexts.

b. That means that it will be necessary to have a "semantic reference frame (Semantic Repository)" carefully developed and held up to date.

c. The maintenance of such a semantic reference frame will consequently make it possible to open the choice of the languages for the e-learning and E-training customers. If not, we move towards the use, by default, of a "BASIC english" and the impoverishment of our linguistic and educational cultures followed-up by the collapse of our educational institutions and professional trainings.

3) How to use Cartago to make visible *multiculturally* and *multilinguistically* concepts which are distinct for national, institutional, cultural and linguistic reasons? For example the school level that marks the beginning and the completion of obligatory teaching then the realms leading to the long studies (university and different) and those conducing to professional carriers.

a. The answer to this question will pass by a rigorous and elaborated organization of the conceptual level of the Cartago model, in fact of the "level terminological entry (conceptual entry) of the TMF" and complementary structures that we propose in the following Cartago model:

(1) A "semantic reference frame (Semantic Repository)", based upon:
(2) Referential Thesaurus,
(3) Registers of proper names and their reference frames of standardized transliterations,
(4) Editorial reference frames of the sources indicated a priori according to Dublin Core meta-data.

b. It is then important to set up Groups of people and resources structures taking charges each a section language. More the languages are spoken in different NBs more these groups must manage in common their terminological resources to reveal some of the semantic distinctions as the only guarantee of real interoperability issue. Languages belonging to various NBs: it is the case of course of English, French, Spanish, Arabic, but also for example of Swedish present at the same time in Sweden and Finland. It will be less the case of Korean and Finnish.

c. Let's insist on the fact that there is not reason to entrust a language section to a NB because the ultimate objective is to describe, in a given language, the whole of the training proposals or modes of apprehension of e-learning technologies and associated research work with an interoperability objective. In other words, that means that any contributor of Cartago must be sensitive to expressions in other languages and for other cultural contexts of his national educational, cultural and institutional characteristics: this deeply induces impact on the prosperity of its national and linguistic area.

d. Under these "languages sections" are located the "term sections" in which terms considered to be synonyms in the language of the concept of the terminological entry are collected. The correspondence between the terms of various languages is established on the level of the concept. Certain syntactic differences are brought by categories of grammatical data (gender, numbers,

topic, part of speech…), the semantic differences are really specified only in the semantic reference frame (Semantic Repository).

4) Cartago and the languages sections: some precisions

    a. Basic assumptions

        i. En and fr which are located at the level of the section *terminological entry* (concept) are has to distinguish well from the languages sections on TMF level of the language section. It is totally forbidden, in the section *terminological entry* to define terms. As a corollary any introduction of term *fr* or *en* into the term section will have to be preceded by reformulation or repetition by the concept in the language section.

        ii. In other words, it must be clear that bilingualism en & fr of the conceptual level of Cartago are collective responsibility for the experts of all the languages and cultures and that structural characteristics of education or training in Australia, GB, IR the USA etc or those of Quebec, Ontario, FR, CH, YOU etc…. must be managed (even for these 2 languages en & fr on the level of Semantic Repository

5) Cartago and the common management of the co-operation of the various educational and linguistic cultures. Description GENETRIX: development environment.

    a. Basic assumptions: Cartago is basically designed to allow a rigorous and qualified management of various levels of drafting and editorial responsibility. An environment of work in conformity with the terminological standards ISO TC37 is proposed by GENETRIX on which was installation a DTD and a transformation XLS in conformity with TMF model Cartago. This tool allows via the access rights management, a first level of multipartners and multilingual co-operative work. This environment is currently an instrument of test. Its finalization and its perpetuation are in hand. A description of the environment as of the access conditions can be obtained by the intermediary of the project team on the address: fabulameda@mshparisnord.org.

    b) In short, it seems necessary to define between the Cartago partners a *modus operandi* guaranteeing a harmonious functioning of the co-operation and the synergy of the contributions. It will be necessary to know to preserve the multipolarity while sharing the same XML model, the same semantic reference frame and the interfaces adaptable to various writings and languages

    c) The text ISO/IEC JTC1 SC36 WG1 N0123 "Document to provide a starting point for ISO/IEC 2382-36 (2nd Edition)" contains 59 terminological entries related   to   e-learning.
For demonstration of what could be a collaborative work about terminology, these entries have been encoded in XML (conforming to the Genetrix format as defined in ISO 16642 Annex C restricted to the elements found in ISO/IEC

2382 standards). These entries have been loaded on a web server.

This text gives information about:

i. How to query this database
ii. How to edit existing entries and how to create new ones
iii. How to add one language to existing entries (the application is UNICODE compliant)
iv. How to extend the ISO/IEC model in order to cope with cross-cultural concept and lexical dissimilarities

6) The IsoGen2382 online databank

- **_How to query the online database extracted from SC36 WG1 N0123_**

  - Connect to http://www.genetrix.org
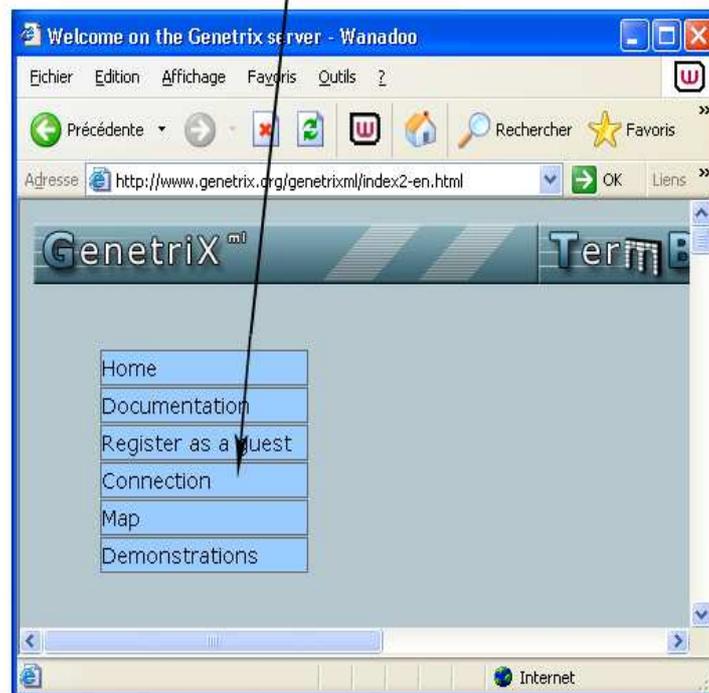
Figure 1 - Connection
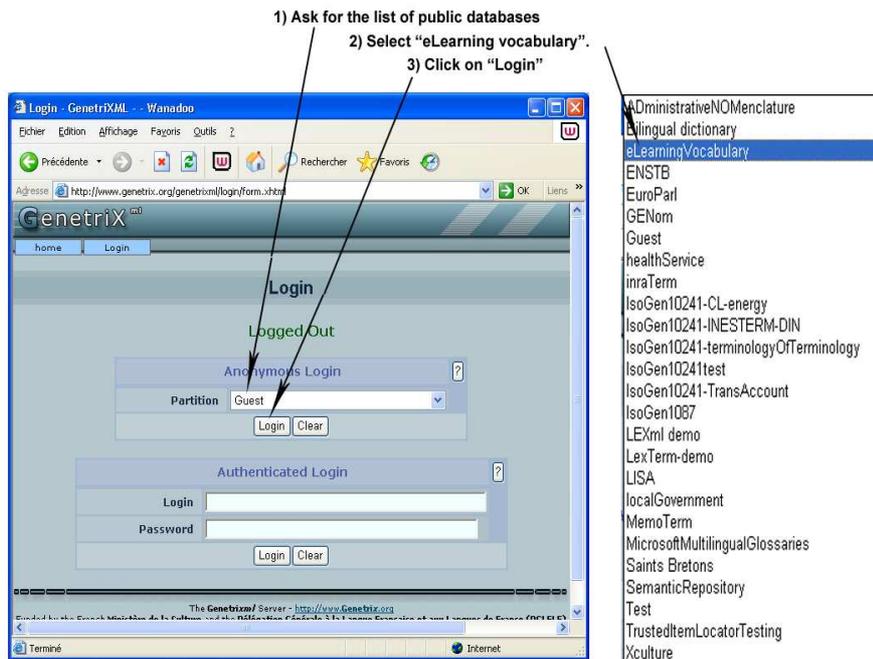
Figure 2 – Login

Select « **Query** » and then « **Linguistic** »
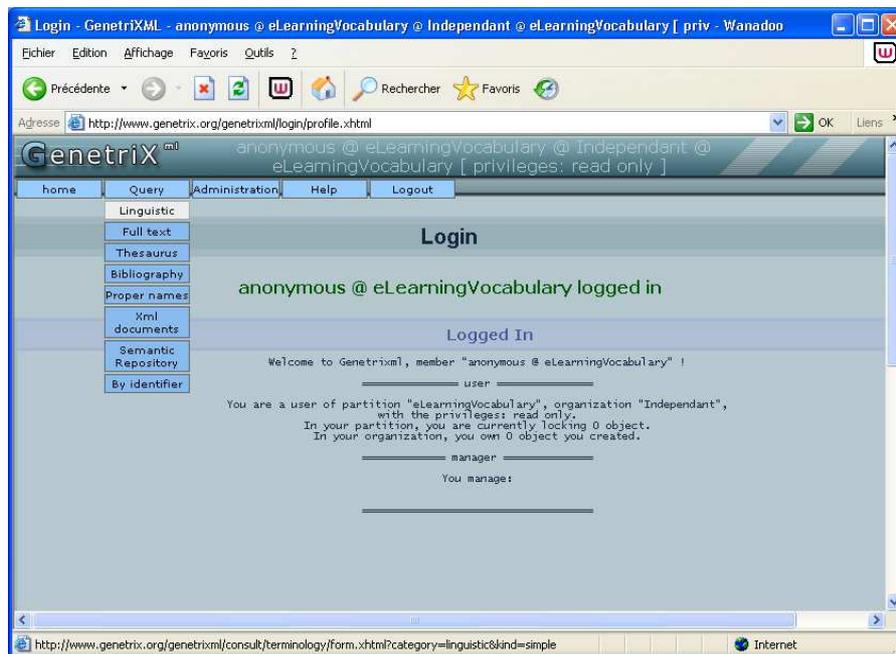
Figure 3 - Choosing a linguistic query

- Type a term (you can use stars for truncation)
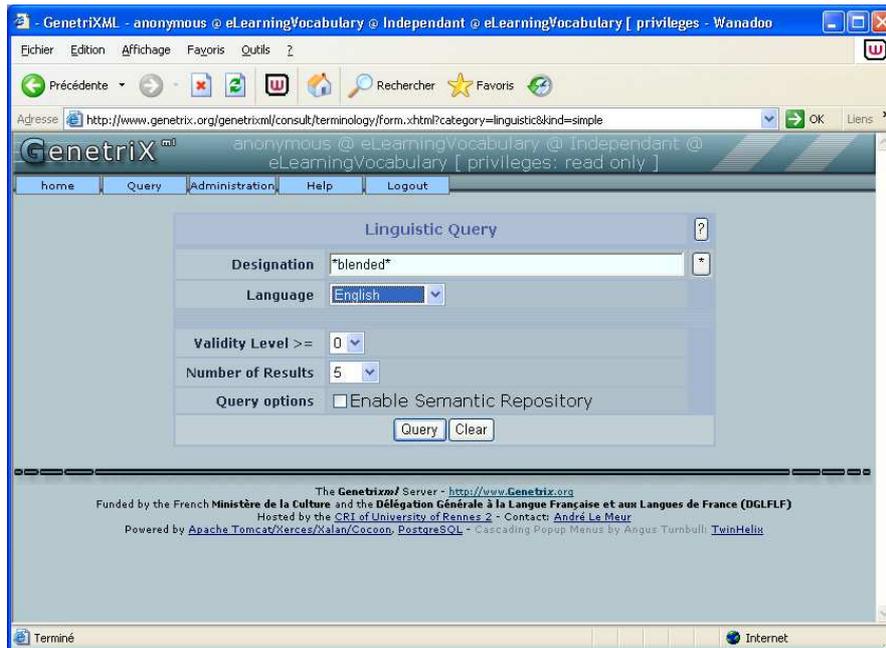- Select a language
- Click on "**Query**"

Figure 4 - Formulating a query
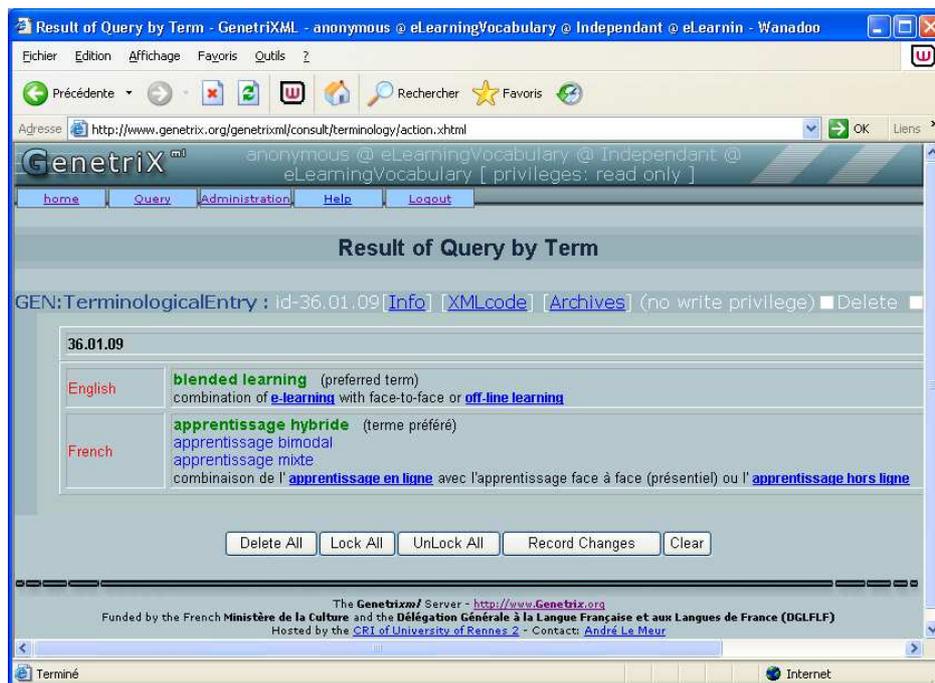
…there should be an answer…

Figure 5 - Display of an entry

(Note that blue-underscored terms are links pointing to the entry where these terms are defined)

- ***Adding one language to the database by creation of a transitory language entry***

For adding a new language to an existing entry there are two possibilities: by modifying the entry in an XML editor as seen before or by editing a special structure called "transitory language entry" that can be sent to the server who will add this language to its target entry.
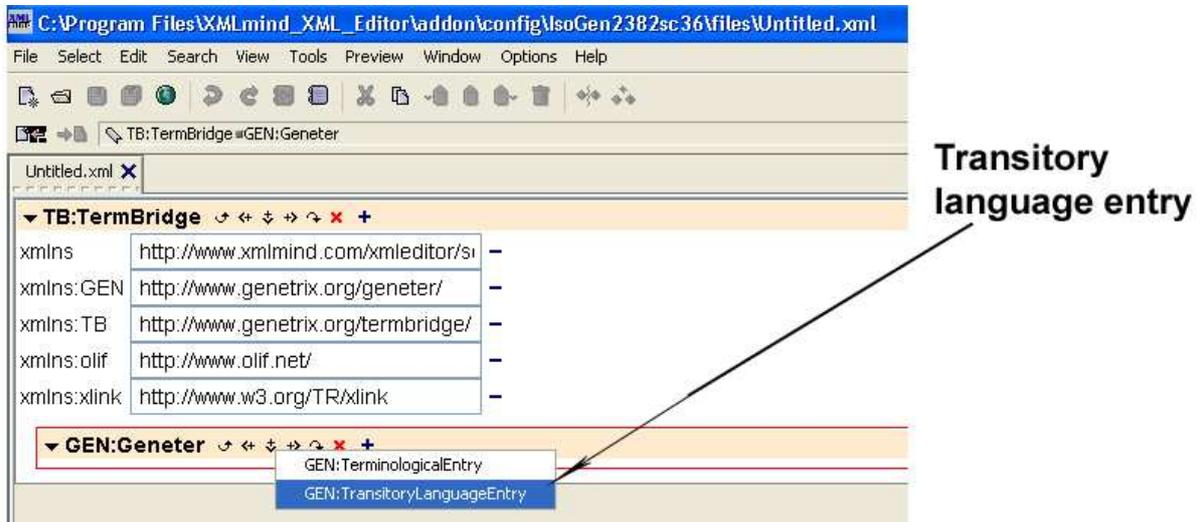

Fig 6 - Choosing New transitory language entry

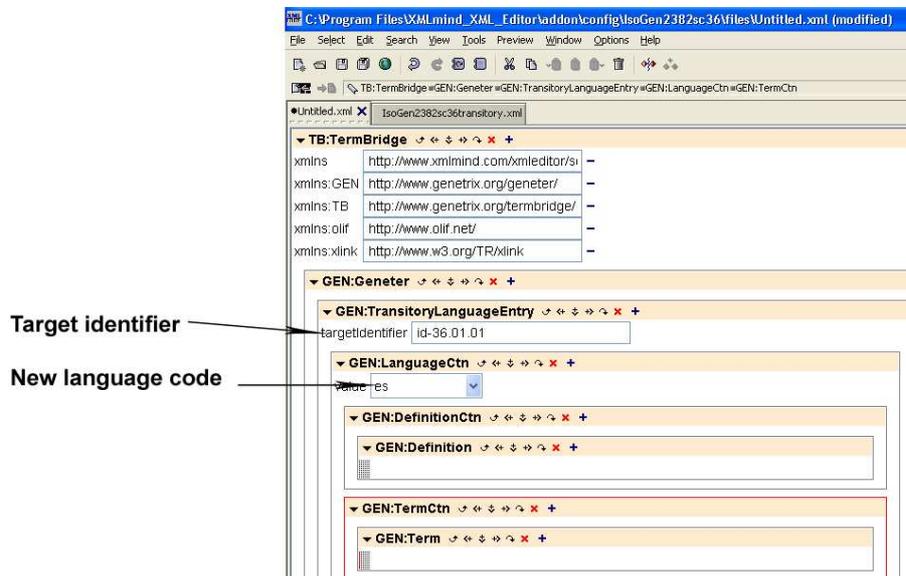The following screen shows how to add the Spanish (language code "es") to the first entry (id-36.01.01).

Fig 7 - Filling a transitory entry

i.By filling an empty pattern

The file IsoGen2382sc36**files\IsoGen2382sc36transitory.xml** contains an empty pattern (derived automatically from **IsoGen2382sc36\files\IsoGen2382sc36.xml**) for all the entries of the existing e-learning data base.

In this pattern, target identifiers are filled and note contains the English and French part of the original entry (in order to help the editor).

You can open it (**File**/**Open**), fill all the language section for all the entries and then load it to the Genetrix server
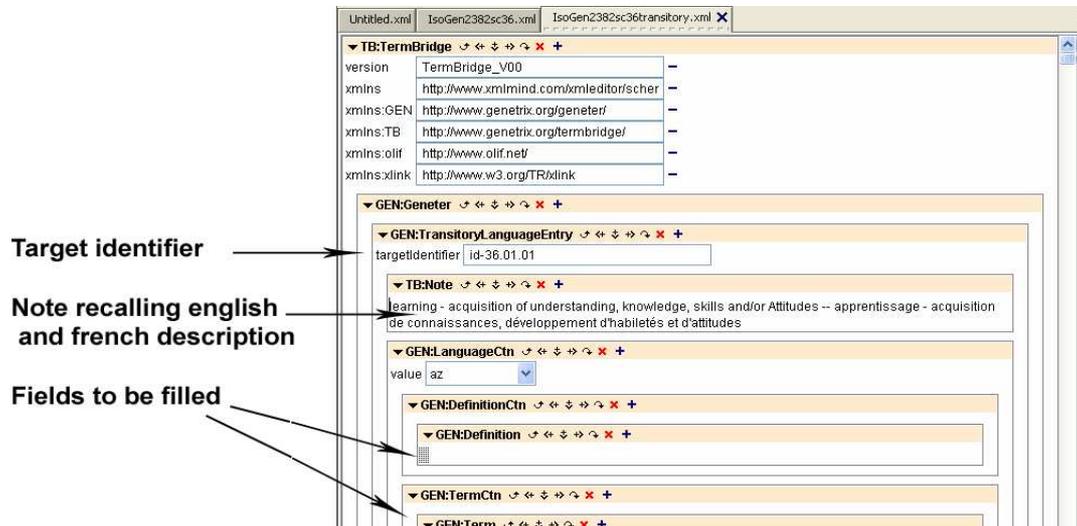

Fig 8 - A pre-filled transitory entry

When such a file has been completed, it can be sent to the Genetrix server that will add this language section to all the entries.

ii.Generating a new empty pattern

After new entries have been added in a terminological file, It can be useful to generate a new empty pattern for transitory entries. This can be done by running an XSL style sheet with four parameters

- the XSL style sheet name
- the source file name
- the option -o followed by the target file
- the attribute name codLang followed by the language code of each transitory entry