Multilingual Electronic Dictionaries for Cross Language IR[1]

A. Vaquero[a], F. Sáenz[b], and A. Barco[c]

[a] Departamento de Sistemas Informáticos y Programación. UCM. Madrid. Spain.

e-mail: avaquero@sip.ucm.es

[b] Departamento de Arquitectura de Computadores y Automática. UCM. Madrid. Spain.

http://babel.dacya.ucm.es/usuarios/fernan/fernan.html

[c] Computing Consultant. Madrid. Spain

e-mail: abarco@teleline.es

Abstract

A sound methodology for the development stages of multilingual electronic dictionaries is defined. In particular, the proposed methodology can be applied to build more restricted electronic dictionaries, i.e., one of Informatics in Spanish, as is the case of our current project. The proposed methodology relies on two basements. The first one is a good classification of the terms. The second one is a conceptual model of the linguistic objects, which are present in a dictionary, and all the relations among them. This model is independent from the language, and can be extended to any set of languages. Thus, the reached model is adequate for implementing multilingual dictionaries needed to build cross language IR systems. We also present the foreseeable evolution of our project and the implications related to the lexical knowledge bases which suit the right characteristics to be used as knowledge handlers in cross language IR.

1 Introduction

Currently much textual WWW information is handled in English, but this situation is changing. Moreover, the users generally would prefer to query in their native language. Undoubtedly, handling information from the Web is a major challenge demanding multilingual solutions. Cross language IR involves multilingual electronic resources, dictionaries and ontologies, among others. Here we are focused on multilingual electronic dictionaries.

The kind of dictionaries needed for CLIR is very different from the currently available dictionaries which generally have been built very dependently on a language.

The method and the model we proposed are absolutely general and, as we reach them by developing an electronic dictionary of Informatics in Spanish, we think the best way to understand the general case is to begin from the particular ones. Why Informatics?

Informatics becomes more pervasive in our society each day. Habits and languages are continuously modified due to this process. Language is the most important cultural legacy of a community. Two issues strongly impact culture and language: the new jargons and the use of computers.

Informatics must respect the mother tongue of each community and its effect ought to be positive, i.e., the new terms and new assignments to the old ones ought to be technically and linguistically correct. Thus, a terminological standardisation in Informatics must be established in each linguistic community.

The technical terms to be introduced in a language constitute a matter of permanent discussion. Introducing a new technical term, will could first ascertain whether there is an equivalent word in the existing language. If not, we could create a new term. In this case, typically the term already exists in a language, generally English. Then, the issue becomes, is it better to use the original technical term directly or to

[1] With the support of the Spanish Ministerio de Educación y Cultura (Programa de estancias de investigadores españoles en centros extranjeros). Computing Research Laboratory, New Mexico State University at Las Cruces.

consider the different translations from language to language?

The technical languages in the Spanish-speaking community have a strong influence of the English language. But the French influence on computer terms in Spain is strong. Thus, there are many linguistic discrepancies between the different geographical regions of the Spanish community.

There have been few Informatics dictionaries in Spanish. Some are personal projects, more or less fortunate in their achievements. For example [Vaq85,ATI]. The institutional efforts have not been significant [IRANOR78]. We must highlight the works in this field due to Real Academia Española [RAE92] and Real Academia de Ciencias Exactas, Físicas y Naturales [RACEFN90], although, we must claim for more support for this kind of work, since up to now is notoriously insufficient. In this situation, a quality Spanish dictionary is necessary.

Nowadays, such a dictionary must be electronic and on line accessible, characteristics that provide the functionality that an end user would require.

We propose actions and methods not previously employed to answer the above-mentioned questions and to achieve the goal of having a complete set of standard terms in the Spanish community. For each community the actions must be based on the coordination of two kinds of high level technical specialists, say computer scientists and linguists. For achieving optimal results, a sound methodology must be established, based on an appropriate classification of the different source fields, which provide the Informatics term. To facilitate this objective, we have developed a first outline for classifying the Informatics terms. This is the first step in developing a multilingual dictionary, taking into account the previous considerations. The lack of such a dictionary is strongly felt in the Informatics community.

This paper is organised as follows. In Section 2, we set the objectives desired for our project. In Section 3, we explain the needed methodology to accomplish this task, and restrict ourselves to a limited methodology due to our scarce resources. In Section 4, we define a Informatics term classification as a fundamental study for a balanced, homogeneous, and functional dictionary. In Section 5, we define the conceptual model of the data base for an electronic dictionary which supports the proposed ideas. In Section 6, we explain the current state and the evolution of the project. Finally, in Section 7, we summarise the conclusions of this work and hints for future work.

2 Objectives

At long term, we envision our electronic dictionary integrated in an adequate multilingual knowledge base , with semantic representation of the lexicon .For this, it is necessary in the first place to develop the terminological data base with the sufficient completeness and robustness. Once the terminological data base has been built, we must integrate it into the chosen knowledge base. So, we must develop our dictionary coherently to the knowledge base structure, as we explain later in Section 5.Once the integration has been a fact, we can take benefit from the applications already developed for this knowledge base, particularly machine translation.So we may develop natural language processing applications in the computing field, therefore improving the power of the knowledge base. In this way, it may be easy to get a product line of on-line or printed dictionaries and other on-line linguistic resources able to be applied directly to CLIR.

Next, we explain the methodology developed for the terminological data base creation.

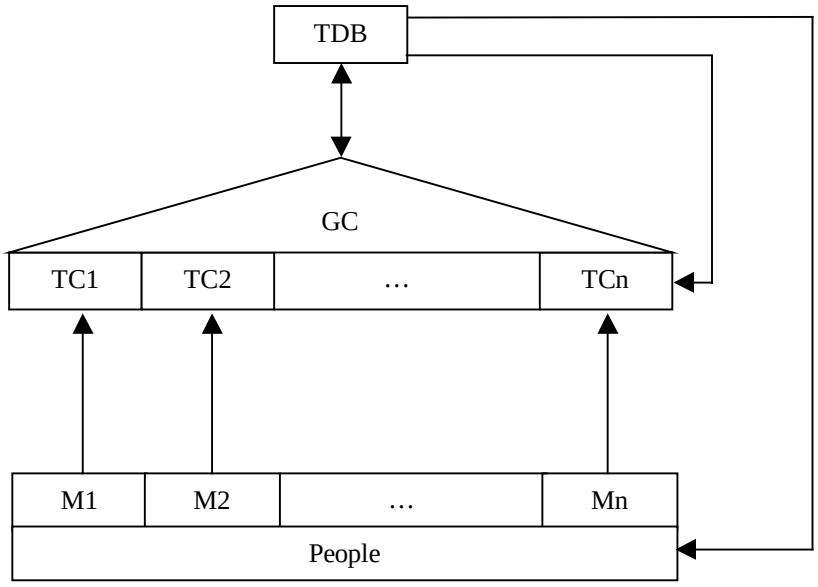
3 Methodology

Firstly, the linguistics sources must be correct. For example, the biggest source of computer terms is English, and it is convenient to review it, because terms as "compiler" are incorrect, and frequently, in Spanish, the translations are simply literal ("compilador"). The desirable final result would be a multilingual Informatics dictionary with the Spanish as the target language.

For achieving such a dictionary it is necessary a very exigent project. It has to be considered as own by the Spanish-speaking community.

Next, we pose the conditions that such a project must accomplish. First, it is necessary to establish a technical category taxonomy for organising the contents of the dictionary. Second, the terminological proposals and the control of the project must depend on coordinated informatics and linguists. And third, the operational method has to ensure the conditions to be met as well as the task progress and the quality of the project.

The objective is to develop a terminological data base with a methodology which takes into account the active participation of the community interested in the project through a cooperative work method. Figure 3-1 shows the methodology we propose. In this figure, we are concerned with Informatics, although the project is aimed to deal with every topic of the language. Current information technologies are very adequate to implement this methodology, in particular the use of Internet as a fast and reliable communication platform.



TDB: Terminological Data Base
GC: Global Committee
TCi: Technical Committee
Mi : Matter

Figure 3-1 Proposed Methodology

A small team, such as the one dealing with this project, composed of a team director and several team members (one linguist and computer scientists), with scarce resources can accomplish the first condition for the project, as we explain in the next section. But the others require resources that are out of our control. However, it is necessary to begin working on this direction with the available resources.

The methodology for building a terminological data base system consists of three main tasks: First, gathering information; its goal is to provide all the needed information for the electronic dictionary, i.e., terms, term translations, and term definitions. Second, building the terminological data base in order to be able to support the gathered information. Finally, developing an application interface for the end user, which is not considered in the rest of the paper.

Up to now, we have worked mainly in the first task, the one of the outmost importance from a linguistic point of view. Its activities are the following: looking for information sources, gathering interesting information, and providing correct translations and definitions under our criteria. These criteria are the result of the experience in technical translations for McGraw-Hill of about fifty people during twenty years. For example, all the Microsoft standards in Spanish have been set by this team, although certain recommendations have not been followed in the human-machine interfaces due to commercial reasons.

The first activity is organised by the team director, who assigns different search topics for each team member. Therefore, each member has to look for information sources all over both the on line and printed worlds in its search topic (different glossaries, dictionaries, forums, ... related to the Informatics world).

The second activity, gathering interesting information, relies on the responsibility of each team

member, who will also have to store the interesting information. Since the team director controls the inclusion of each term in the dictionary, only rejection of terms should be accomplished when no doubt is, otherwise it has to be stored.

We will accomplish the last activity, providing correct translations and definitions, in two pipelined stages. In the first place, we gather the information from selected sources in an intermediate data base. This data base, called Store (Almacén) will contain the information as it is, together with other information fields which will be explained later. In the second place, we provide definitive information derived from the Store and our experience, storing it in a data base called Work (Trabajo). This dynamic data base (in the sense it will be augmented and upgraded) is the source for the electronic data base, which concerns to the last two tasks.

We propose a methodology for carrying out these two stages. The methodology corresponding to the first stage is depicted in Figure 3-2. This figure shows each team member (TM*i* , $i \in \{1,..,n\}$), who gathers and stores information in a particular data base (StoreTM*i*). TMx collects and prepares all of the StoreTM*i* in order to update Store, the unique public data base for browsing. TMy broadcasts Store to all the members. The processing protocol is as follows: each time a member TM*i* prepares a data base StoreTM*i* with enough (under a predefined criterium) contents, she/he sends it to TMx. When a sending is performed, an empty data base StoreTM*i* is selected for gathering new information and repeating the procedure. TMx, in turn, prepares as soon as possible the new information in order to be broadcasted by TMy to all the team members.

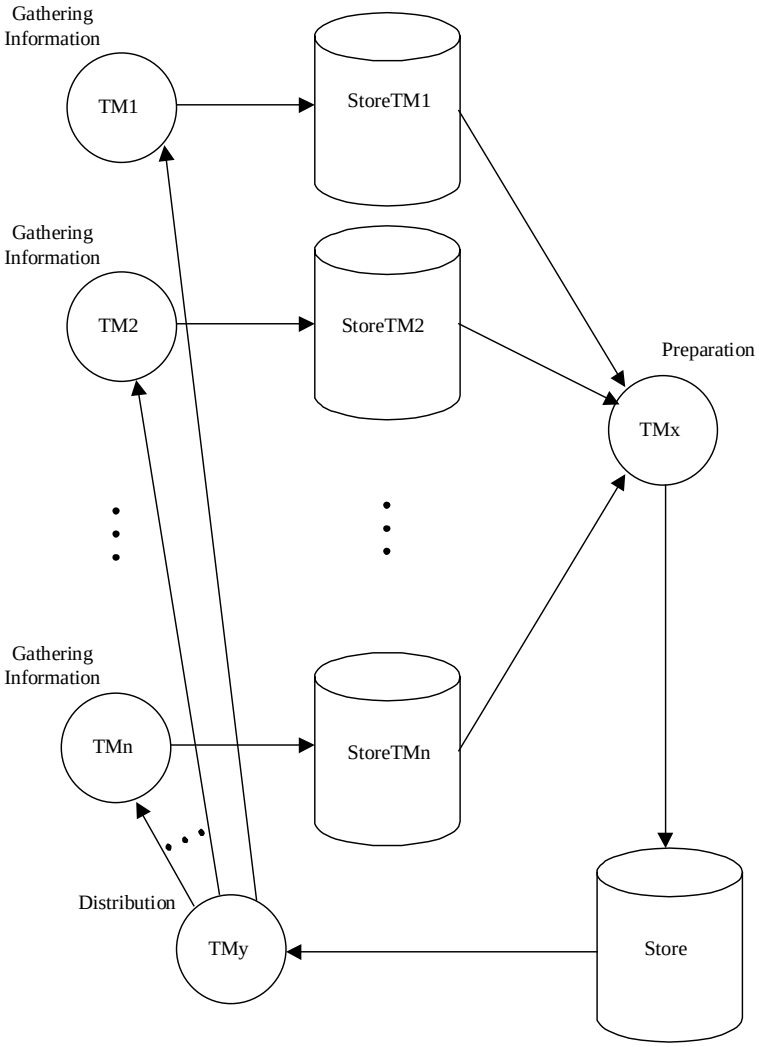


Figure 3-2 Providing correct translations and definitions, First pipelined stage

The methodology corresponding to the second stage of the last activity, by which we build the data base Work, is shown in Figure 3-3. From the information in the data base Store, each team member TM*i* adds entries to an empty data base called WorkTM*i*. Each team member is responsible for classifying each term under its criterium and identifying itself as owner for later refinements. TM*i* sends WorkTM*i* to TD, the team director, who supervises and modifies the information received and updates the data base Work. Finally, TMy broadcasts Work to all the team members. This procedure contains two cycles:

- Entries supervised by TD which form part of the data base Work are accorded when each team member sees the modifications due to TD. If someone disagree, they formulate new proposal in WorkTM*i*, therefore performing the cycle.

  In order to effectively carry out this procedure, it is needed to record the added terms when adding entries to WorkTM*i*, along to record the entries in Work modified by TD. So, it is possible for each team member TM*i* to identify those entries which have been modified by TD and that belongs to TM*i*. Moreover, since TD may reclassify an entry, looking for modifications in Work have also be guided by the term classification, and, so, a team member who has not added a given reclassified entry could automatically detect it.

  Under this procedure, it is also allowed that several team members may propose modifications to entries that do not correspond them by its category, so that non-experts in the field of the given category can provide another point of view.

  This methodology is part of a seminar stage of the project, and it is intended for avoiding redundant work disallowing several team members to generate duplicated entries. Nonetheless, a given term can be provided by several team members under different meanings and categories. In a further stage, all the terms will be reviewed by all the team members.

- The second cycle comes from the fact that it is possible that some terms in Store are not recognised as one's own by any team member. It is needed to detect the absents from Store by contrasting Work with Store, which is carried out by the team member TMz. This cycle will be achieved later in the project, when a high degree of accord had been reached for the most entries.

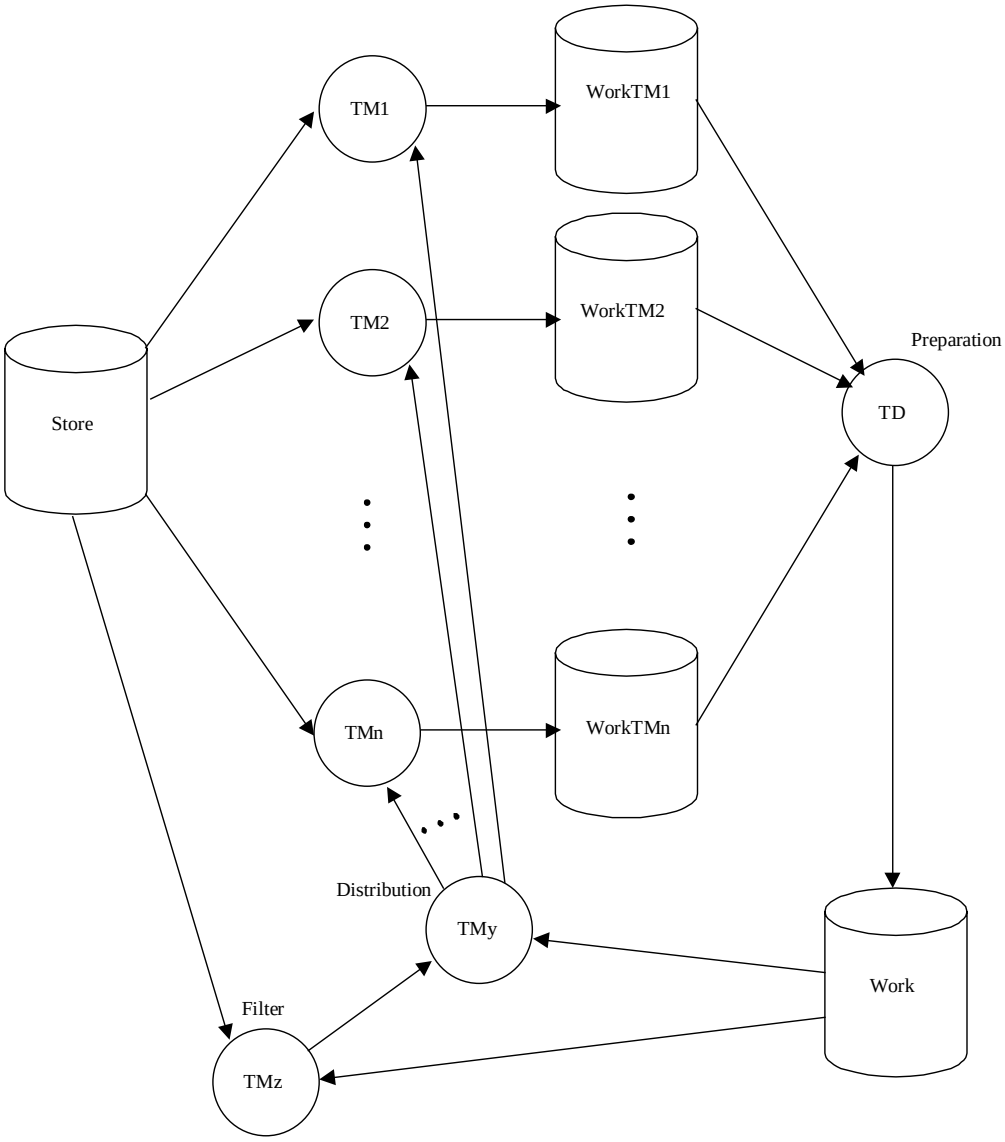


Figure 3-3 Providing correct translations and definitions, Second pipelined stage

4 Meaning Classification

Our aims in classifying meanings are several. First, to provide a useful facility for an electronic dictionary. So, meanings embody additional semantics which provide more information to the end user. The system may also gain a new dimension because it is possible to generate in an automatic way specialised dictionaries under different categories (Computer Networks, Fundamentals, ...). Second, to ensure a balanced dictionary by adding enough terms from different fields. Having the terms classified it is easy to check how many entries are under a given category. Third, to provide homogeneity to the dictionary: each team member is specialist in a given computation field and she/he is the adequate for providing the information related with meanings classified under her/his knowledge area. Moreover, we will count on opinions of external people (specialists, linguists) for the meanings in which they are experts. The needed procedure of collecting the classified meanings for sending to specialists is simplified providing classification to meanings. Moreover, classification of meanings is important to integrate the terminological data base into the multilingual knowledge base.

4.1 Classification Table

In this section we show the classification table we have developed and adopted for meaning classification, which is an upgrade of [Vaq92]. This table is not intended for classifying knowledge areas into subjects as a curriculum does; instead, it tries to group thematically the meanings according to its use in the computing contexts of users, developers, industrials and academics. This table is the result of our current thought, and is intended to show the guidelines for its upgrading by example. Table 4-1 shows the classification both in English and Spanish (left and right hand side, respectively). Each entry has a usual hierarchy notation with space-dotted numbers.

1. Other Science Foundations	1. Bases de otras ciencias
1.1. Mathematics	1.1. Matemáticas
1.1.1. Logic	1.1.1. Lógica
1.1.2. Algebra	1.1.2. Álgebra
1.1.3. Combinatorial	1.1.3. Combinatoria
1.1.4. Numerical Calculus	1.1.4. Cálculo numérico
1.2. Physics	1.2. Física
1.2.1. Electronics	1.2.1. Electrónica
1.2.2. Optics	1.2.2. Óptica
1.2.3. Mechanics	1.2.3. Mecánica
1.2.4. Quantum Physics	1.2.4. Física cuántica
1.2.5. Electromagnetism	1.2.5. Electromagnetismo
1.3. Philology	1.3. Filología
1.4. Psychology	1.4. Psicología
1.5. Biology	1.5. Biología
2. Computing Fundamentals	2. Fundamentos
2.1. Computability	2.1. Calculabilidad
2.2. Switching	2.2. Conmutación
2.3. Coding	2.3. Codificación
2.4. Automata and Formal Languages	2.4. Autómatas y lenguajes formales
2.5. Information and Communication Theory	2.5. Teoría de la información y de la comunicación
3. Hardware	3. Hardware
3.1. Digital Systems	3.1. Sistemas digitales
3.2. Computer Architecture	3.2. Arquitectura de computadoras
3.2.1. Microprocessors	3.2.1. Microprocesadores
3.2.2. I/O Devices	3.2.2. Unidades e/s
4. Software	4. Software
4.1. Programming	4.1. Programación
4.1.1. Imperative Programming	4.1.1. Programación imperativa
4.1.1. Structured Programming	4.1.1.1. Programación estructurada
4.1.2. Modular Programming	4.1.1.2. Programación modular
4.1.3. Object Oriented Programming	4.1.1.3. Programación orientada a objetos
4.1.2. Declarative Programming	4.1.2. Programación declarativa
4.2.1. Functional Programming	4.1.2.1. Programación funcional
4.2.2. Logic Programming	4.1.2.2. Programación lógica
4.2.3. Constraint Programming	4.1.2.3. Programación con restricciones
4.1.3. Programming Languages	4.1.3. Lenguajes de programación
4.1.4. Data Structures	4.1.4. Estructuras de datos

4.1.5. Algorithms	4.1.5. Algoritmos
4.2. Software Engineering	4.2. Ingeniería del software
4.2.1. Design	4.2.1. Diseño
4.2.2. Production	4.2.2. Construcción
4.2.3. Test and Maintenance	4.2.3. Pruebas y mantenimiento
4.2.4. Development Tools	4.2.4. Entornos de desarrollo
4.3. Data Base Management Systems	4.3. Sistemas de gestión de bases de datos
4.3.1. Data Bases	4.3.1. Bases de datos
4.3.1.1. Relational Data Bases	4.3.1.1. Relacionales
4.3.1.2. Hierarchical Data Bases	4.3.1.2. Jerárquicas
4.3.1.3. Network Data Bases	4.3.1.3. En red
4.3.1.4. Object Oriented Data Bases	4.3.1.4. Orientadas a objetos
4.3.1.5. Deductive Data Bases	4.3.1.5. Deductivas
4.3.2. Data Base Languages	4.3.2. Lenguajes de gestión de bases de datos
4.4. Knowledge Bases	4.4. Bases de conocimientos
4.5. Information Retrieval	4.5. Recuperación de información
5. Computing Systems	5. Sistemas informáticos
5.1. Functions, Characteristics, and System Properties	5.1. Funciones, características y propiedades de sistemas
5.1.1. Security	5.1.1. Seguridad
5.2. Computer Networks	5.2. Teleinformática
5.3. Operating Systems	5.3. Sistemas operativos
5.4. Multimedia	5.4. Multimedia
5.4.1. Computer Graphics	5.4.1. Infografía
5.4.2. Audio Systems	5.4.2. Sistemas de audio
6. Applications	6. Aplicaciones
6.1. Office Computing	6.1. Ofimática
6.1.1. Word Processing	6.1.1. Procesadores de texto
6.1.2. Data Sheets	6.1.2. Hojas de cálculo
6.2. Business Computing	6.2. Informática de gestión
6.2.1. Information Systems	6.2.1. Sistemas de información
6.3. Educational Computing	6.3. Informática educativa
6.4. Legal Computing	6.4. Informática jurídica
6.5. Medical Computing	6.5. Informática médica
6.6. Industrial Computing	6.6. Informática industrial
6.7. Architecture and Town Planning	6.7. Arquitectura y urbanismo

Table 4-1. English and Spanish Classification Table

This table is intended to be dynamic in nature, i.e., classification can be refined and upgraded as new terms are added to the dictionary, and, as well, as specialists suggest new categories.

In order to classify meanings we consider on the one hand that every entry in the hierarchy will be used, no matter its hierarchy level. We must note that every entry in the table is a term itself in the dictionary, so that all categories are non-empty. On the other hand, we will try to avoid the creation of new categories as intersection of several predefined ones by assigning meanings to the most adequate category.

5 The Conceptual Model of the Terminological Data Base

In this section we show the development of the conceptual model for the terminological data base. First, we pose the considerations we have taken into account for developing it. Then, we show the conceptual

model we propose to represent the information in the electronic dictionary.

5.1 Considerations

The mapping of Spanish terms to English terms (i.e., translations) is one of the goals of this project. It is straightforward to notice that an English term can be translated in general into several Spanish terms, and vice versa. The mapping comes from the relations among meanings and terms. For modelling the relationships, it is necessary to determine the relations which define pairs English term - Spanish term through the meanings, which are independent from the language.

In every language there exists the known *naming problem* [KP93], which consists of two facts: one is polysemy, so that a term can have several meanings, and the synonymy, that one meaning can have assigned different terms, as can be observed in Figure 5-1. In this Figure, Term 1 and Term 2 are synonyms and have a shared meaning, as so for Term 2 and Term3, under another meaning. Moreover, Term 2 is polysemic.

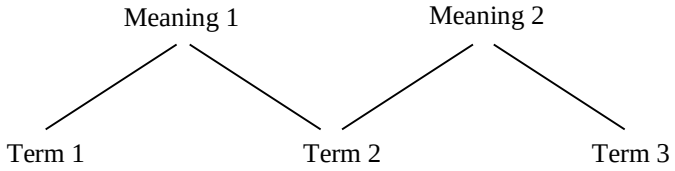


Figure 5-1 Polysemy and Synonymy.

We note some remarks about the relationships between categories, meanings and terms. On the one hand, a given term can belong to several categories under *different* meanings. On the other hand, a given term can belong to several categories under the *same* meaning. In Figure 5-2 we show a term T2 that is assigned to meanings M12 and M21, that respectively belong to categories C1 and C2. We also show the term T that is assigned to meaning M, which belongs to both categories C1 and C2. Polysemy is present in T2, and synonymy is also present in T3, and T4, as it can be seen. T1 is neither polysemic nor synonym. TC1 and TC2 are the terms used to denote categories C1 and C2, respectively.

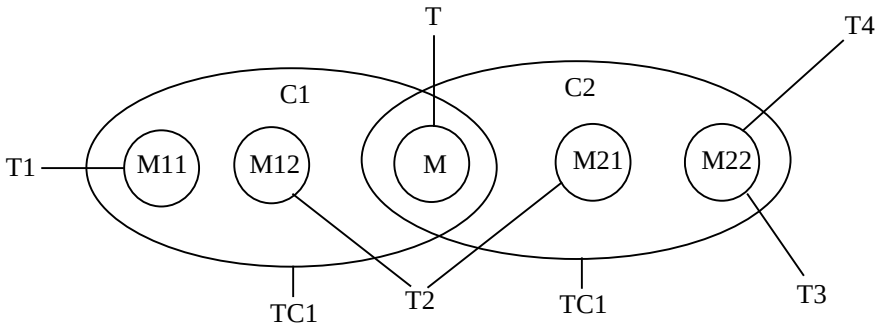


Figure 5-2 Relationships among categories, meanings and terms. Extensional definition.

In Figure 5-2, the set of meanings ({M11, M12, M}) in C1 is the extensional definition of category C1. The same applies to all the categories.
Term classification is mandatory. From the classification table (Table 4-1) and the previous Figure, we have the mechanism needed to assign categories to terms.
We must also note that a category has a meaning described by a definition. Previous figure does not embody this fact. Now, in order to embody the meanings related to categories, we transform the scheme of Figure 5-2 in another equivalent. This is shown in Figure 5-3.

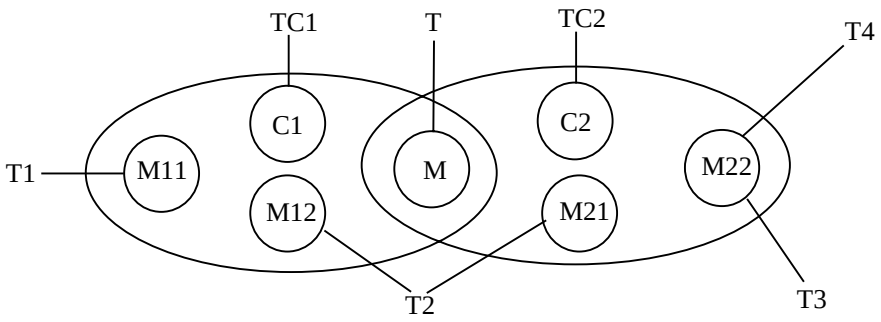


Figure 5-3 Relationships among categories, meanings and terms. Intensional definition.

Now, C1 is the meaning of the category C1, and TC1 is the term assigned to such meaning, and the same applies to C2 and TC2. Then, we have one more meaning in each category. This meaning is the intensional definition of the category.

Because its own importance, term classification emerges as a need, which not only proves useful for the end user but also for the development of the dictionary, when assigning work to experts in each computing field, who can propose correct definitions for meanings, and correct both English and Spanish terms for meanings.

For a given language, we have a set of computing terms that holds the relationships with categories and meanings shown in Figure 5-3. If we now think of several languages, the same applies for each one. Then, relationships between terms from different languages come from considering at the same time the involved schemes.

At this point, we make an important observation. A sound conceptual model is necessary to integrate the terminological data base into the multilingual lexical one. We have in mind lexical knowledge bases based on ontologies (e.g., MikroKosmos [MK]), other than monolingual on line lexical resources (e.g., WordNet [Mill90]). The conceptual model we develop here is coherent with the concept of ontology. Therefore, the implementation of the terminological data base from that conceptual model must facilitate its integration into an ontology based lexical knowledge base.So we can assume the implementations to be built from this model will acomplish the conditions to be used rightly CLIR operations.

5.2 The Entity-Relationship Model

With the considerations just posed we have the entity-relationship model shown in Figure 5-4.

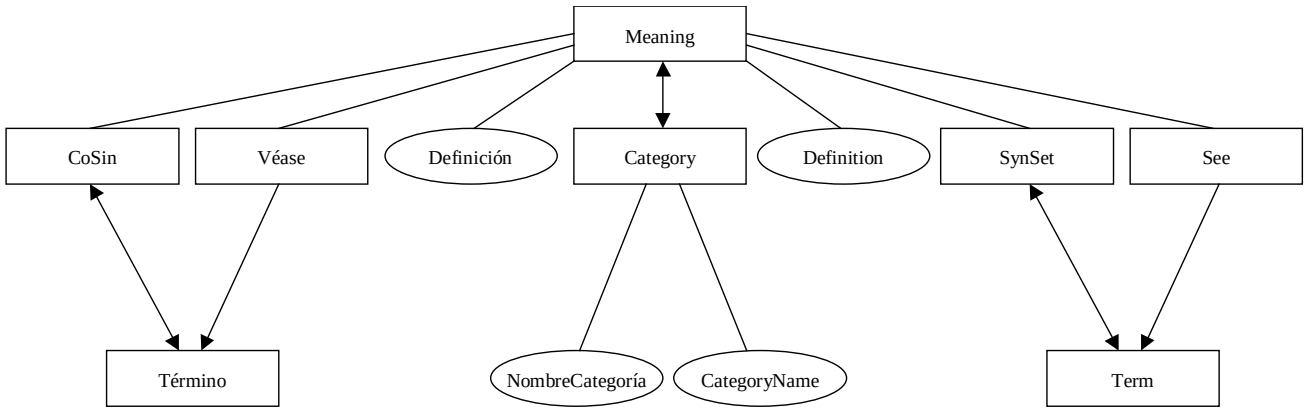


Figure 5-4 Entity-Relationship Model

In this Figure (following some recommendations in [Pre98,SKS98]), entity sets are represented with rectangles, attributes with ellipses, and relationship sets with directed and undirected lines. If B has an

incoming line from A, this denotes a one(A) to many(B) mapping cardinality. Double arrows denote many to many mapping cardinalities. Undirected lines denote one to one mapping cardinalities. Relationship set names (not shown in this Figure) label each line.

We depict in this Figure the entity Meaning, the central entity other entities rest on. The entity SynSet denotes the English synonym set (SynSet - Synonym Set). The relationship set between both entities is one to one. The entity Term represents all the English terms that compose the terminological data base. The relationship set between SynSet and Term is many to many since a synonym set contains several terms, and a term may be contained in several synonym sets (obviously, with different meanings). Figure 5-5 embodies this idea, in which Term 1 and Term 2 are synonyms and has a shared meaning, as so for Term 2 and Term 3, under another meaning. Moreover, Term 2 is polysemic.

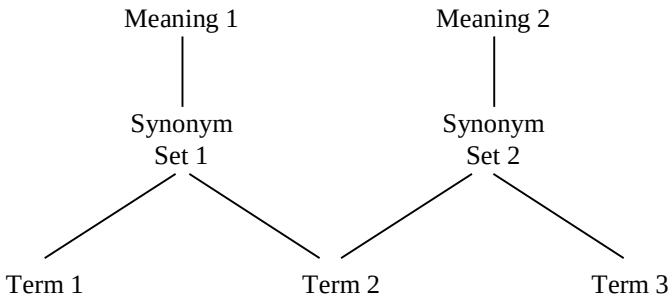


Figure 5-5 Polysemy and Synonymy related with the synonym sets.

The entity See denotes the set of English terms related under a given meaning. The relationship set between Meaning and See is one to one. The relationship set between See and Term is one to many, because a meaning may refer several English terms. The entity Definition represents the textual definition given to a meaning. The relationship set between Meaning and Definition is one to one. The entity Category denotes the category each meaning belongs to. The relationship set between Category and Meaning is many to many since many meanings are in a category, and a meaning can be in several categories (this situation is expected to be reduced to the minimum since our goal is to keep the classification as disjoint as possible). The entity Category has two attributes: CategoryName and NombreCategoria, which correspond to the textual name of the category in each considered language, English and Spanish, respectively. Meaning has two attributes: Definition and Definición, which correspond to the textual definition in the same considered languages. The remaining entities (CoSin, Véase, Término) are homologous to the respective entities (SynSet, See, Term).

The skeleton shown in this section can be straightforwardly applied for a multilingual dictionary, by adding the needed entities and relationships for each added language.

6 Evolution of the Project

The first task of the project is going on. As this task has no end, we hope that it will be minimally useful and usable in one year.

The remaining tasks are going on in parallel. We hope that we will have a friendly human-machine interface also in one year.

An important step is the formalisation of definitions in terms of ontologies. We will start on it just when the knowledge base has been selected.

Next, we envision the integration of the Informatics dictionary in any adequate multilingual lexical knowledge base, including Spanish. This integration will firstly allow to enlarge and improve the standardised lexical resources and then, to extend the machine-translation fields, particularly English-Spanish translations in Informatics. There are a Spanish-English bilingual corpus of high quality available to be used as a source of inspiration to improve the current machine-translation capabilities and, more important, our knowledge of natural languages.

The improved capabilities of the chosen linguistic resources are going to make feasible CLIR in the Web. So, we must explore the role of dictionaries and ontologies in CLIR, i.e., for searching, extracting, summarising, etc, textual documents written in different languages for satisfying queries expressed in native language. The envisioned workplan is impressive.

7 Conclusions

We have developed a classification table and the guidelines for improving it. Its current state permits use it for building an electronic dictionary in a systematic way. This table is intended to be improved along the dictionary building process. This classification is not only useful for users and specialists, but also for the automatic language processing, for building thematic dictionaries, and so on.

We have also developed a conceptual model for representing the linguistic objects and the relations among them. This model allowed us to reach an entity-relationship model, which is necessary for the development of the electronic dictionary as an information system leading to a multilingual lexical knowledge base, as required for CLIR.

References

[ATI] http://www.ati.es/PUBLICACIONES/novatica/glointv2.html

[IRANOR78] "Informática. Vocabulario. Soporte de datos, memorias y dispositivos relacionados". Proyecto de norma española, 1978.

[KP93] B. Katzenberg and P. Piela, "Work Language Analysis and the Naming Problem", Communications of the ACM, Vol. 36, No. 4, June 1993.

[Mill90]G.Miller,"WordNet:An on-line lexical database"Int.Journal of Lexicography,3(4).Special Issue.1990.

[MK] http://crl.nmsu.edu/Research/Projects/mikro/index.html

[Pres98] R.S. Pressman, "Software Engineering. A Practitioner's Approach", McGraw-Hill, 1997.

[RACEFN90] Real Academia de Ciencias Exactas, Físicas y Naturales, "Vocabulario científico y técnico", Ed. Espasa Calpe, 1990.

[RAE92] Real Academia Española, "Diccionario de la lengua española", Ed. Espasa Calpe, 1992.

[SKS98] A. Silberschatz, H.F. Korth, S. Sudarshan, "Data Base System Concepts", WCB/McGraw-Hill, 1997.

[Vaq85] A.Vaquero , "INFORMÁTICA. Glosario de términos y siglas. Diccionario Inglés-Español Español-Inglés", McGraw-Hill, 1985.

[Vaq92] A.Vaquero, "Informática, Educación y Lenguaje". Revista de Enseñanza y Tecnología (ADIE). Nº 7, Junio 1992. pp 11-31.