A Methodology for building Translator-oriented Dictionary Systems.

E. Agirre, X. Arregi, X. Artola, A. Díaz de Ilarraza, K. Sarasola and A. Soroa.

Informatika Fakultatea (Computer Science Faculty)
P.K. 649, 20080 DONOSTIA (Basque Country - Spain)
e-mail: jiparipx@si.ehu.es

Abstract

Interaction between humans and translation tools has been deeply studied in the field of machine-assisted translation. However, support tools for translation are often designed without the co-operation of human translators. The underlying idea is that human translators must adapt to the new technologies, and, it seems that new computerised tools would not need to consider translators' practical uses and experience. On the contrary, we argue that it is worthwhile and necessary to analyse the behaviour of translators in order to fit the assistance tools to their needs.

This paper presents an experiment to incorporate human translators' expertise into an already constructed lexical system. We focus on the design methodology that could be applied for the improvement of other similar tools.

1. Introduction.

Lexical knowledge bases (LKBs) are a common support for natural language processing tasks. Commonly, the LKBs are of general purpose, that is, the information they contain is intended to be used in several applications. Therefore, they can be seen as reusable resources.

Nonetheless, it is well known that these bases intended to be generic hardly adapt to the specific features of each application. The gap between the generic knowledge and the specific applications is usually solved *ad hoc*. The adaptation needed implies to deal with a wide variety of aspects, such as those concerned with the representation and implementation of lexicons or their usage.

In this paper, we will emphasise on the operational or functional aspects and we will present a methodology for adapting a lexical knowledge base to a specific task. This adapting methodology consists of three steps: (1) specification of the real work environment, (2) elicitation of the functional knowledge and, (3) incorporation of the elicited knowledge into the dictionary system.

In relation to the first step, we consider that it is necessary **to specify the real work environment** because no good further developing is possible without a suitable functional specification (a basic software-engineering principle). Actually, we are interested in reusing the lexical knowledge as the basis of a dictionary system for humans when translating words.

The second step deals with the **elicitation of the functional knowledge**. In our opinion, as tools for translation can not be satisfactorily designed without the co-operation of human translators, any attempt to incorporate task-dependent behaviour into a dictionary system should begin with a study of the tasks involved and the real users' interaction with the dictionary. That leads us to work with human translators in a real context, and to extract expertise knowledge from their activity.

In the third step, we adopt a task-oriented methodology in order **to incorporate the elicited knowledge into the dictionary system**. The foundations of this process lie on KADS[1] ("Knowledge Acquisition and Documentation Structuring") (Schreiber *et al.,* 93). Among other features, KADS assists in clearly distinguishing levels of knowledge.

The structure of this paper is as follows: in section 2, the structure of the initial lexical knowledge base is presented. Sections 3, 4 and 5 describe the three steps of the methodology. Finally, in section 6 some conclusions are presented.

2. The initial Lexical Knowledge Base and the Multilingual Environment

In the Lexical Knowledge Base (LKB) that constitutes the basis of this work (Agirre *et al.,* 94c) a typed frame-based model has been adopted. In such a model, lexical concepts are represented by frames, which are interrelated by slots representing lexical-semantic relations. The data of the LKB have been extracted automatically from machine-readable dictionaries (Agirre *et al.,* 94b). In that way, the proposed model allows us to represent different dictionaries of different languages. The design of the LKB is based on the following principles:

---

[1] This methodology has been succesfully applied in the design of knowledge-based systems. But, as far as we know, it has not been used in the design of machine-assisted translation systems.

a) Relational representation model. Starting from the analysis of definitions of a French dictionary we extracted different types of relations (Artola, 93) such as: synonymy and antonymy, taxonomic relations as hypernymy/hyponymy —obtained from definitions of type "genus et differentia"—, meronymy, and others.

b) Typed information. Different types of objects have been defined and hierarchically organised. The representation proposed for them captures the common features that will be inherited. Fundamentally, these types of objects represent attributes related to the lexical and semantic information. These descriptions constitute the meta-knowledge of the LKB.

c) Capability to make inferences. Using this capability, implicit information contained in the dictionary will become explicit.

2.1 Description of the Monolingual Knowledge Base

The knowledge representation schema chosen for the monolingual LKB is composed of three elements, each of them structured as a different knowledge base (figure 1):

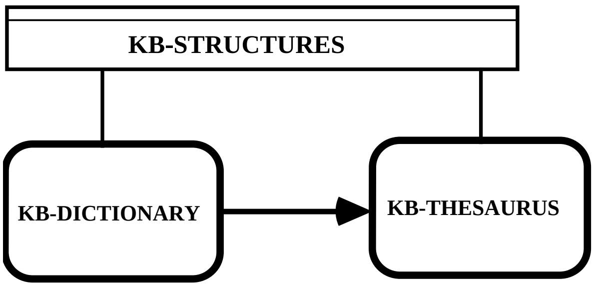


Fig. 1. The components of each Monolingual Knowledge Base.

- KB-THESAURUS is the representation of the dictionary as a semantic network of frames, where each frame represents a *one-word concept* (word sense) or a *phrasal concept*. Phrasal concepts represent phrase structures associated to the occurrence of concepts in meaning definitions. Frames —or units— are interrelated by slots representing lexical-semantic relations.

- KB-DICTIONARY allows access from the dictionary word level to the corresponding concept level in the LKB. Units in this knowledge base represent the entries (words) of the dictionary and are directly linked to their corresponding senses in KB-THESAURUS.

- KB-STRUCTURES contains meta-knowledge about concepts and relations in KB-DICTIONARY and KB-THESAURUS: all the different structures in the LKB are here defined specifying the corresponding slots and describing the slots by means of facets that specify their value ranges, inheritance modes, etc. Units in KB-THESAURUS and KB-DICTIONARY are subclasses or instances of classes defined in KB-STRUCTURES.

The subclasses defined under KB-STRUCTURES are the following:

- ENTRIES, it groups dictionary entries belonging to KB-DICTIONARY.
- DEFINITIONS, which groups word senses classified according to their POS.
- REFERENCES, concepts created in KB-THESAURUS due to their occurrence in definitions of other concepts ("definitionless").
- CONCEPTS, that groups, under a conceptual point of view, word senses and other conceptual units of KB-THESAURUS.

The classification of conceptual units under this last class is as follows:

- **TYPE-CONCEPTS** correspond to Quillian's "type nodes" (Quillian, 68); this class is, in fact, like a superclass under which every concept of KB-THESAURUS is placed. It is further subdivided into the classes ENTITIES, ACTIONS/EVENTS, QUALITIES and STATES, which classify different types of concepts.
- **PHRASAL-CONCEPTS** is a class that includes concepts similar to Quillian's "tokens" —occurrences of type concepts in the definition sentences—. Phrasal concepts are the representation of phrase structures that are composed by several concepts with semantic content. A phrasal concept is always built as a subclass of the class that represents its head (the noun of a noun phrase, the verb of a verb phrase, and so on), and integrated in the conceptual taxonomy. Phrasal concepts are classified into NOMINALS, VERBALS, ADJECTIVALS, and ADVERBIALS.
- **AMBIGUOUS-CONCEPTS**, concepts that, after the analysis phase, are not yet completely disambiguated (lexical ambiguity).

The links between units in KB-THESAURUS and KB-DICTIONARY are implemented by means of slots tagged with the name of the link they represent. These slots are defined in the different classes of KB-STRUCTURES.

The representation model used in the system is made up of two levels:

- ***Definitory level***, where the surface representation of the definition of each sense is made. Morphosyntactic features like verb mood, tense; noun definiteness, etc. are represented by means of facets attached to the attributes.
- ***Relational level***, that reflects the relational view of the lexicon. It supports the deductive behaviour of the system and is made up of *relational attributes*, which may eventually contain deduced knowledge. These attributes, defined in the class TYPE-CONCEPTS, are the implementation of the interconceptual relations: ANTONYMY, AGENT, TYPE-OF, etc.

Figure 2 gives a partial view of the three knowledge bases that form the monolingual LKB with their correspondent units and their inter/intra relationships.

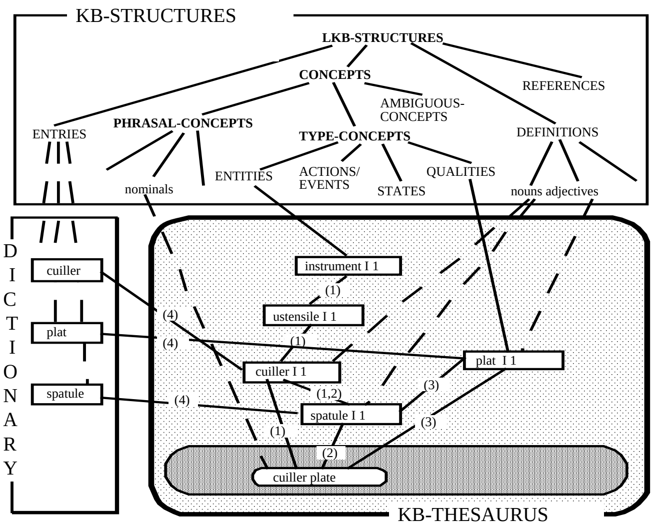


Fig.2. The French monolingual Dictionary Knowledge Base.

—— SUBCLASS link
- - - _ MEMBER-OF link (instance)
(1) _ Taxonomic Relation: HYPERNYM/HYPONYM
(2) _ Specific (meta-linguistic) relation: KIND-OF/KIND-OF+INV
(3) _ PROPERTY/PROPERTY+INV relation
(4) _ ENTRY-WORD/WORD-SENSE relation

In KB-THESAURUS, some of the links representing lexical-semantic relations are created when building the initial version of the knowledge base, while others are deduced later by means of specially conceived deduction mechanisms e.g. deduction of inverse relationships, taxonomy formation, etc.

When a dictionary entry like ***spatule I 1:*** *sorte de cuiller plate* (*spatula: a kind of flat spoon*) is treated, new concept units are created in KB-THESAURUS (and subsidiarily in KB-DICTIONARY) and linked to others previously contained in it. Due to the effect of these links, new values for some properties are propagated through the resulting taxonomy.

In the example, although it is not explicit in the definition, *spatule* is "a kind of" *utensil* and so it will inherit some of its characteristics (depending upon the inheritance role of each attribute). Fig. 2 also shows the types of concepts used: *spatule I 1* and *cuiller I 1* are noun definitions which will be considered subclasses of ENTITIES while *plat I 1* (an adjective) is a subclass of QUALITIES. The phrasal concept unit representing the noun phrase *cuiller plate* is treated as a hyponym of its nuclear concept (*cuiller I 1*).

2.2 The Bilingual Environment

In the bilingual environment, two monolingual knowledge bases are related by means of a bilingual one. The monolingual modules follow the model described above. Next we will describe the bilingual module in which links among the corresponding concepts of each monolingual environment will be established. Note that the definitions of concepts of the monolingual environments do not change when they are integrated into the bilingual environment.

The KB-S/BILINGUAL knowledge base includes the definition of the classes and attributes needed in the representation of the bilingual dictionary. Three different classes have been defined:

a) Source-Unit Class: Defines the type of the link between the unit of the bilingual dictionary and its corresponding concept of the Monolingual Dictionary. It is represented by the *equivalence* relation.

b) Target-Unit Class: The same as the previous but referred to target units.

c) Bilingual Unit Class: It characterises the information about the equivalence relation itself. Different attributes are used in this model in order to enrich the bilingual links between concepts of different languages. Some of them follow:

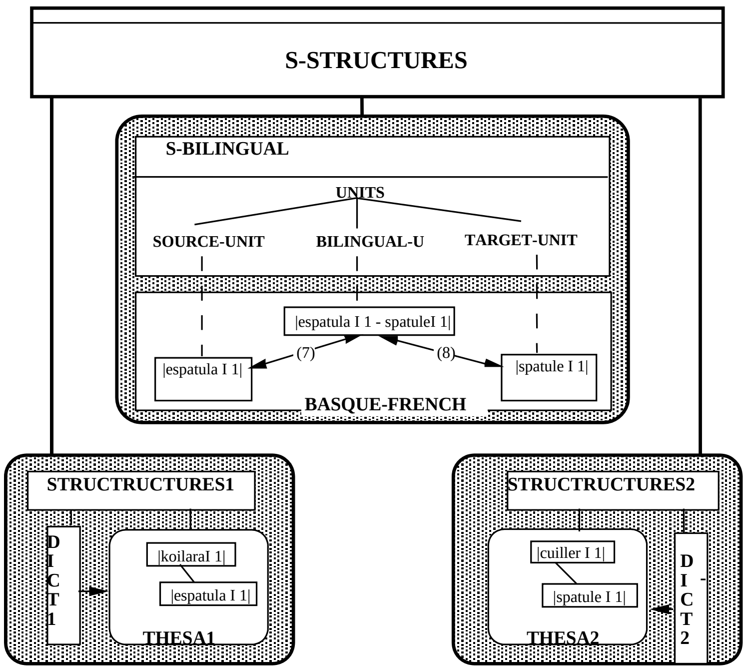
- Level-of-equivalence: it represents the level of equivalence between the linked units. Three different levels of equivalence are proposed to characterise these links. One level represents that the concept of the source language is more general than its corresponding concept in the dictionary of the target language. A second level represents greater specificity, and, finally, the third level expresses conceptual equivalence.

- Type-of-equivalence: the model of representation proposed permits us to establish two types of equivalence. Those that relate two concepts are named conceptual and those that relate concepts with phrasal concepts are named syntagmatic.

In the current version KB-S/BILINGUAL represents the information contained in a Basque/French bilingual dictionary.

When a bilingual dictionary entry like ***spatule:*** *a kind of flat spoon* is treated, a new unit is created and the corresponding links with the monolingual dictionaries (source and target) are established by means of the *EQUIVALENCE* and the *REPRESENTATIVE+INV* relations.

An example of the bilingual environment follows (see figure 3). The figure shows the general organisation of the global LKB in which the correspondences between the monolingual and bilingual environments have been established.



(7) EQUIVALENCE/EQUIVALENCE+INV Relation
(8) REPRESENTATIVE/REPRESENTATIVE+INV Relation

Fig. 3. General view of the lexical knowledge base.

3. Functional view

The lexical knowledge base described (Agirre *et al.*, 94c) provides various access possibilities to data. Even so, limitations are present when trying to exploit this knowledge in a lexical translation context. The cause of this limited usability is that the lexical organisation was thought from a general perspective, without taking into consideration functional aspects. Incorporating this functionality means, in our case, transforming such a lexical knowledge base in a user-oriented dictionary system.

In fact, it involves a task-oriented approach, so the lexical knowledge must be enriched with reasoning mechanisms analogous to those used by humans when they consult a dictionary. A way of adapting the dictionary system could be to incorporate into it both lexical knowledge and knowledge about the use of dictionaries when dealing with a lexical problem. This way, the dictionary system would become an active tool able to better support the task of lexical translation, rather than being only a repository, more or less structured, of words and definitions.

In our opinion, the idea of the "active" dictionary introduced by Martin W. and Al B.P.F. (1990) is particularly relevant: "...the use of a dictionary can be seen as a typical problem-solving activity, and user-orientation should involve both (static) knowledge and dynamic features (strategies, aims, needs) of the intended user". Even considering dictionaries as human user-oriented tools (traditional concept of the dictionary), they could incorporate "dynamic features" by means of appropriate computational functions.

This notion of the dictionary as a dynamic tool is adequate to be applied in the context of machine-assisted translation. Such a dictionary should be useful either when understanding source lexical items, when searching for equivalents, or even when predicting target lexical forms. For that, it would need to incorporate reasoning mechanisms in order to exploit the explicit and implicit information from the dictionary. Furthermore, along with the usual information about the meaning of the entries, dictionaries should show how to use words in context. In other words, we advocate that dictionaries should actively co-operate in finding the correct translation.

Let us bind this functional approach by specifying our idea of a dictionary tool: *The dictionary is a problem-solving lexical tool that receives as input a translation context and a specific function to be applied. The output is the result of the application of such a specific function in the given context.*

For instance, if the human translator needed to lexicalise (lexical problem) the idea of *an hymenopterous insect with sting* (translation context), s/he would try a kind of thesaurus-based search (specific function) into the dictionary system whose result might be *wasp* (output).

4. Elicitation of functional knowledge

In order to deal with the functional approach we need to specify the behaviour of the dictionary by describing the functions that could be executed on it. The definition of these functions must determine the context of use and the parameters (input and output data) of each of them.

Certainly, it would be desirable to have a well-founded theory about the use of a dictionary and its interaction with the human users. The use of dictionaries has been previously researched from different perspectives, (Hartman, 85; Atkins & Knowles, 90; Nuccorini, 94).

Along with these studies, some attempts to model the use of dictionaries in translation have been already carried out. Namely, in a research study made in the *School of Translation and Interpreting* of Maastricht (Starren & Thelen, 90), the use of dictionaries is organised in four steps: discovering meaning, finding receptor language equivalents, checking meaning of receptor language item, and formulation of final translation.

Different models of the translation process are presented in (Sager, 94). These models exhibit certain limitations that make them unsuitable for our specific purpose, such as:

- They are made in a statically descriptive way which idealises the process. No indication is given on the nature or complexity of the mental processes involved.
- They are theoretically speculative and not based on empirical data.
- They do not account for task specifications.

One of the conclusions we extracted relies on the fact that the use of the dictionary must be seen as a process, rather than as a specific action. In (Neubach & Cohen, 88) this idea is encouraged. Therefore, no good approach to the use of dictionaries can be made without analysing this process.

In order to face this kind of analysis, we have adopted an empirical way. Our study has not been limited to a questionnaire-based method to collect information. As an alternative, we have used both direct observation and personal interviews as presented below:

a) Direct observation protocols. The translators were given several texts to be translated (in our case French and Basque texts) along with several dictionaries (monolingual and bilingual with different characteristics) in order to record their problems, the solutions they adopted, and the tasks they carried out. The aim was to characterise the activity of human translators by observing the task of translating words, expressions, context-dependent phrases and even paragraphs (rarely). Each time the human translator used a dictionary, the unit to be translated, the dictionary used, the consulted dictionary entry and the type of consultation were recorded.

b) Personal interviews with professional translators. These interviews have allowed us to detect different uses of the dictionary according to their experience in the subject. Additional questions were posed to the experts: what characteristics a dictionary should have in order to be useful when translating, their interest about having computerised dictionaries and their expected functionality, and so on.

The data extracted led us to interesting considerations (some of them are already well-known): *i)* expert translators and occasional ones need distinct and adapted help, *ii)* some translators, mainly occasional ones, find bilingual dictionaries very useful, *iii)* multi-word terms are a source of failure when using dictionaries, *iv)* context is important when translating a text, *v)* dictionaries for translation must give grammatical and usage information, *vi)* the proximity between languages is helpful, but attention must be paid to "false friends"; dictionaries must prevent translation errors derived from them.

4.1 General description of the dictionary-use model.

Starting from what we have observed and recorded the model of the use of dictionaries when translating words has been formalised into three steps i.e., conceptualisation, specification and operationalisation.

The conceptualisation step was limited to distinguish and classify the entities involved in the space of the model. We have classified the entities into three types: objects/roles, subtasks and states.

The specification has consisted on clearly describing the entities and their relations. It is a task-oriented specification, given that tasks can be seen as the core entities where objects and states are described and related.

Finally, the operationalisation has dealt with the way of carrying out the specified tasks. In this step we were concerned with the strategy of the translator, and therefore we have used an algorithmic language.

The model of expertise obtained in that way is organised in a top-down structure. The top level of such a structure (see Fig. 4) shows the generic tasks, and leaves of the hierarchy express the specific ones.

In our model, we consider tasks as divided into composite (expressed by uppercase in figures) and primitive (lowercase in figures). All the inner nodes express composite tasks.

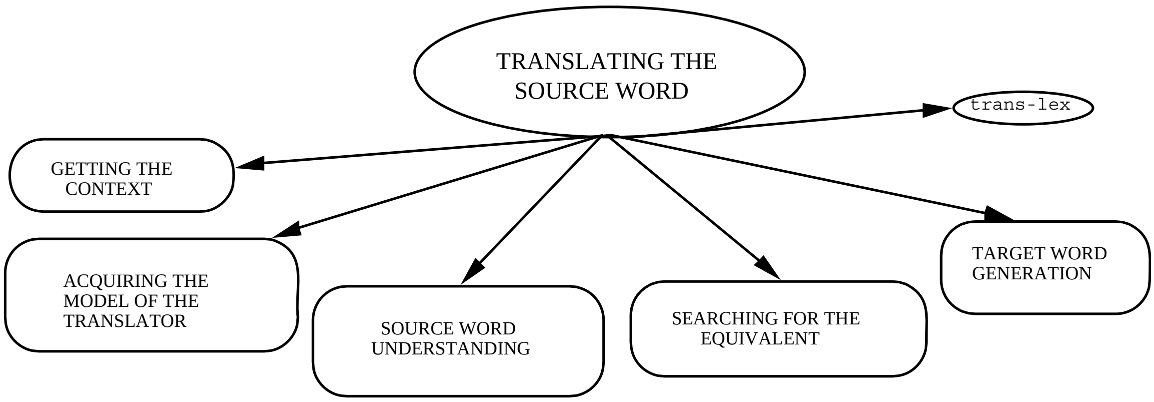


Fig. 4. First level of the decomposition diagram of the tasks involved in the lexical translation process.

The composite tasks are described as non-trivial processes to be decomposed into subtasks when they are carried out. At the same time, each of these subtasks can be decomposed into other subtasks. The process continues until the total decomposition into primitive tasks.

Primitive functions refer to the basic uses of the dictionary identified as useful for translators when translating words. The primitive functions constitute the procedural knowledge (methods in the sense of object-orientation) associated to the entities of the dictionaries. Twenty four primitive functions have been identified and specified. A list of them is shown in the appendix 1.

For example, the *rths* (thesaurus-like search of concepts) action is one of the twenty four primitive functions. This action performs the search for lexical units, based on some constraints, in case the user has an imprecise idea about the exact concept s/he is looking for. The primitive task *rths* would be preferentially used when verifying the meaning of a source word or when finding production hypotheses, which are its parent-tasks.

We will illustrate all these ideas with an example; let us concentrate in the *TARGET_WORD_GENERATION* composite task (last subtask to be performed in order to solve the main task, in figure 4). It would be carried out to get a suitable word that corresponds to the source word to be inserted in the context; in other words, its output is the target lexical word-form corresponding to a (pre-lexical) meaning, translation of the source lexical concept (figure 5). Some prerequisites must be fulfilled to activate this task: *i)* there is a source concept to be translated and an equivalent concept in the target language, *ii)* the meaning and the morphological information associated to the source concept and the context in which it has appeared are well-known, and, *iv)* besides, the system knows the translator's profile.

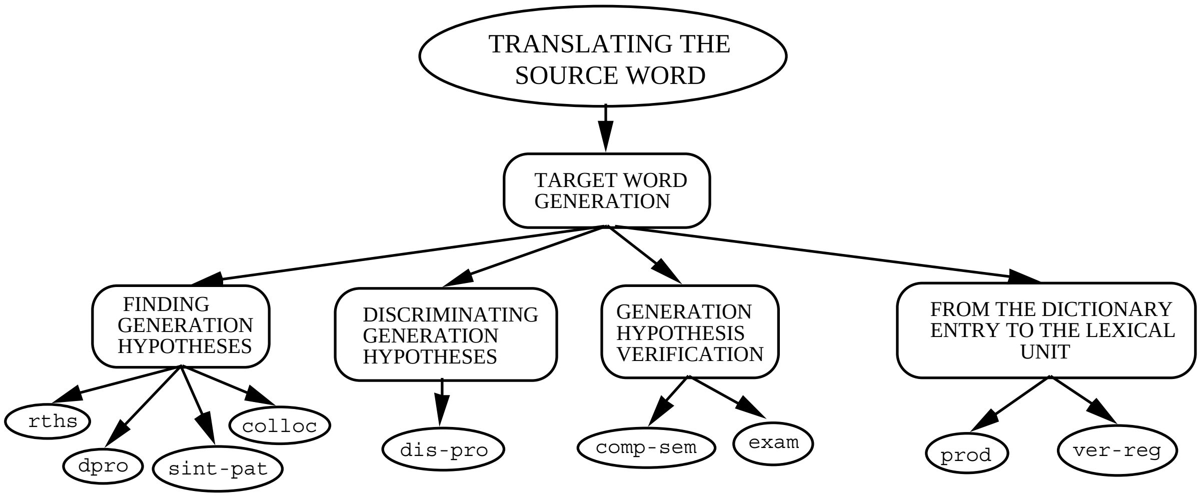


Fig. 5. Decomposition diagram of the *TARGET_WORD_GENERATION* task involved in the lexical translation process.

In our empirical model, four subtasks are involved in the resolution of the task: *FINDING_GENERATION_HYPOTHESES*, the goal of which is to produce the target word-form, considering that the translator has enough knowledge about the translation of the source concept; *DISCRIMINATING_GENERATION_HYPOTHESES*, which focuses on discriminating the possible productions of the target concept looking at their properties; *GENERATION_HYPOTHESIS_VERIFICATION*, which verifies whether the final hypothesis for the target concept is correct in the context in which the source concept has appeared, and the *FROM_THE_DICTIONARY_ENTRY_TO_THE _LEXICAL_UNIT* subtask, the objective of which is to produce the final word-form to be used as the translation of the source word-form. These subtasks are expressed in terms of other subtasks, which in this case are primitive functions.

5. Integration of the functions into the lexical knowledge base

The set of primitive functions constitutes the interface between the human user and the lexical knowledge base. From this view, we consider the dictionary as an abstract object in which the lexical knowledge is the data, and the set of primitive functions constitutes the functional and operational layer.

The definition of the functions involves their detection and identification (introduced in the previous point) and their specification. In order to face the specification of these functions, we adopt CML (Conceptual Modelling Language), which is one of the languages used in the CommonKADS methodology (Schreiber *et al.* 94).

Following the KADS philosophy, each task, either composite or primitive, has a descriptive frame in the task structure. The *TARGET_WORD_PRODUCTION* task, for instance, is specified as follows:

```
task-knowledge
  task#36:        target_word_generation
    task-specification
        task-definition
          goal:  to get a word corresponding to the source word and
                 suitable to be inserted into the context.
          Input: target_concept: concept
                 target_definition: definition
          output:target_unit: lexical_unit
        task-body
          type:  composite
          parents:      translating_the_source_word
          sub-tasks:
                        finding_generation_hypothesis,
                        discriminating_generation_hypothesis,
                        generation_hypothesis_verification,
                        from_the_dictionary_entry_to_the_lexical_unit
          additional-roles:
                        context: text_model,
                        translator: translator_model,
                        languages: language*,
                        list_of_productions: concept*,
                        ?verified: boolean,
                        target_entry: dictionary_entry,
                        target_morphology: morphology
```

```
acceptance-conditions[2]:
              (available_concept(target_concept)) or
              (available_definition(target_definition))
competence-conditions[3]:
              (produced_unit(target_unit))
control-structure: -- algorithm for carrying out the task
```

Primitive tasks are specified in a similar way. Below we present the description of the *rths (thesaurus-like search of concepts)* primitive task:

```
task-knowledge
  task#24:       rths
    task-specification
        task-definition
          goal:  to find the concepts that satisfy a given set of
                 constraints
          input: constraint_expression:  expression,
                 preferred_dictionary: dictionary,
                 result_language: language
          output:list_of_concepts: concept*
        task-body
          type:  primitive
          parents:     meaning_verification,
                 finding_production_hypothesis
          sub-tasks:   nil
          additional-roles:
                 result_concepts_variable: variable
```

5.1 An overview of the primitive functions.

As we have already mentioned, primitive tasks refer to the basic actions carried out by translators when using dictionaries. Therefore, they become procedural units that can be directly executed by the system.

These primitive tasks are classified according to the composite tasks in which they occur.

Source text understanding

The definition request *(ddef)* can be considered as the core function in the word understanding task. It takes as input a concept, an explanatory-level, a dictionary and a language, giving as output a definition. The following example is a definition query for the meaning of `guêpe` *(wasp)* with *inherited* as explanatory-level. The result is the textual definition along with other extra information (“inherited”) that, not being explicit in the dictionary (in italics in the example), is deduced.

---

2 The acceptance field states which conditions should a task hold to be executable.

3 The competence field states which conditions will be held once a task has been executed.

```
User.-          DDEF (|guêpe I 1|, inherited, French, ?D)
System.-        Wasp is an articulated hymenopterous insect with
                sting and legs, a bumblebee is a wasp, and a wasp's
                nest has wasps.
```

Other functions related to this task are: reformulation of a definition *(rdef)*, definition verification *(vdef)*, request for properties of a concept *(dpro)*, verification of properties of a concept *(vpro)*, request for differences between two concepts *(ddif)*, request for relationships between two concepts *(drap)*, verification of relationships between two concepts *(vrap)*, thesaurus-like search of concepts *(rths)*, morphological analysis of a word form *(analy)* and request for examples *(exam)*.

Searching for the equivalent

The dictionary system offers a set of primitive tasks in order to accomplish the complex task of searching for a suitable equivalent. The search for potential translation equivalents *(equiv)*, the search for syntactic constructions that correspond to a given pattern *(pat_sint)* and the semantic compatibility between concepts according to a given relation *(comp_sem)* are among the most relevant.

Target unit production

In this task are involved the thesaurus-like search of concepts *(rths)*, the lexical collocation *(colloc)*, the lexical form production *(prod)*, and the verb-regime *(reg-verb)*.

This is an example of use of *rths*:

```
User.-      RTHS((and   (?X   HYPERONYME    |consumer I 1|)
                  (?X   AGENT  |feu I 1|)),
                  Basque, ?X, ?LC)
            Comment: The user asks for verbs in Basque for 'to
            consume' with agent 'fire'
System.-    LC=(|izeki I 1|, |kiskali I 1| )
            Comment: to burn, to blacken.
```

The effective use of these primitive functions involves the enrichment of the initial LKB with translation-oriented knowledge, such as morphological, syntactic, and usage information. The needed adaptation has been carried out adding new attributes and relations to our representation model.

Furthermore, conceiving the dictionary-system integrated into a translation workbench implies the design of the user interface that compiles all the interactions that occur when making use of the dictionary system.

A prototype of the dictionary called MLDS (MultiLingual Dictionary System) has been developed from these premises.

6. Conclusions and future work

In this work we mark the distinction between lexical systems and dictionary systems. Obviously, the lexicon is the basis of any dictionary, but the dictionary is not only a repository of words and definitions.

We refer to the dictionary system as an assistance tool for translators. Starting from that assumption, we develop a methodology to endow the lexical system with the functionality needed by human translators.

Such a methodology has been formalised into three steps i.e., the specification of the real work environment, the elicitation of the functional knowledge, and the incorporation into the dictionary system of the elicited knowledge. The expertise knowledge is formalised by means of made-to-measure primitive functions. They should reflect what a translator might use when consulting any dictionary. So, the dictionary system that compiles all of them is being integrated into a human translator's workbench.

The MLDS prototype has been implemented as a result of this process. In the future, the dictionary-system will be validated and evaluated by human translators.

7. Acknowledgements

The Basque Government and the University of the Basque Country have supported this research (UPV 141.226-TA073/96 and PI96/86 projects).

8. References

Agirre E., Arregi X., Artola X., Díaz de Ilarraza A., Evrard F., Sarasola K.. Analysing word-level translation activity to design a computerised dictionary help system, (Panel Session) *EURALEX´94 Congress*, Amsterdam. 1994a.

Agirre E., Arregi X., Artola X., Díaz de Ilarraza A., Evrard F., Sarasola K.. IDHS, MLDS: Towards Dictionary Help Systems for Human Users, *Semantics and Pragmatics of Natural Language: Logical and Computational Aspects,* Sara, 167-188, 1993a.

Agirre E., Arregi X., Artola X., Díaz de Ilarraza A., Evrard F., Sarasola K.. Sistema diccionarial multilingüe: aproximación funcional, *Boletín SEPLN* , 14, 313-333, 1993b.

Agirre E., Arregi X., Artola X., Díaz de Ilarraza A., Sarasola K.. A methodology for the extraction of semantic knowledge from dictionaries using phrasal patterns, *Proc. IBERAMIA'94* (Caracas), 263-270, 1994b.

Agirre E., Arregi X., Artola X., Díaz de Ilarraza A., Sarasola K.. Lexical Knowledge Representation in an Intelligent Dictionary Help System, *Proceedings COLING'94* (Kyoto), 544-550, 1994c.

Agirre E., Arregi X., Artola X., Díaz de Ilarraza A., Sarasola K., Soroa, A.. Constructing and intelligent dictionary help system, *Natural Language Engineering 2 (3),* 229-252, Cambridge University Press, 1996.

Arregi X.. *ANHITZ: Itzulpenean laguntzeko hiztegi-sistema eleanitza / Sistema diccionarial multilingüe de ayuda en la traducción*, Ph.D. Thesis. UPV-EHU. 1995.

Artola X., Evrard F.. Dictionnaire intelligente d'aide à la compréhension, *Actas IV Congreso International EURALEX'90* (Benalmádena), 45-57. Barcelona: Bibliograf, 1992.

Artola X.. *HIZTSUA: Hiztegi-sistema urgazle adimendunaren sorkuntza eta eraikuntza / Conception et construction d´un système intelligent d´aide dictionnariale (SIAD),* Ph.D. Thesis. UPV-EHU. 1993.

Atkins B.T., Knowles F.E.. Interim report on the EURALEX/AILA Research Project into Dictionary use, in T. Magay, J. Zigány eds., 381-392, *Proceedings BudaLEX´88*. Budapest: akadémiai kiadó, 1990.

Breuker J. (ed.). *EUROHELP. Developing Intelligent Help Systems.* Report on the P280 ESPRIT Project EUROHELP. Kopenhague, 1990.

Fillmore C.J., Atkins T.S.. Starting where the Dictionaries Stop: The Challenge of Corpus Lexicography, in B.T.S. Atkins, A. Zampolli eds., 349-393, *Computational Approaches to the Lexicon.* New York: Oxford University Press, 1994.

Hartman R.R.K.. Four perspectives on dictionary use: a critical review of research methods. The Dictionary and the Language Learner, in A.Cowie ed., 11-28, *Papers from the EURALEX Seminar (Leeds), Lexicographica, Series maior,* 17, Tübingen; Niemeyer. 1985.

Hauenschild C.. The Lexicon and the General Design of MT Systems, *Machine Translation and the Lexicon, Workshop of the European Association for Machine Translation,* Heidelberg, 1993.

Hutchins W.J., Somers H.L.. *An Introduction to Machine Translation.* Cambridge: Academic Press, 1992.

Isabelle P.. Bi-Textual Aids for Translators, Proc.8th Annual Conf. of the UW Centre for the New OED and Text Research (Waterloo), 76-89, 1992.

Johnson R.L., Whitelock P.. Machine translation as an expert task, in S. Nirenburg ed., 136-144, *Machine Translation: Theoretical and methodological issues.* Cambridge: Cambridge Univ. Press, 1987.

Kay M.. *The Proper Place of Men and Machines in Translation,* CSL-80-11, Xerox PARC, 1980.

Krause J., Mittermaier E., Hirschmann A.. The Intelligent Help System COMFOHELP. Towards a Solution of the Practicality Problem for User Modeling and Adaptive Systems, *User Modeling and User-Adapted Interaction,* no. 3, 249-282, 1993.

Martin W. and Al B.P.F.. User-orientation in dictionaries: 9 propositions, in T. Magay, J. Zigány eds., 393-399, *Proceedings BudaLEX´88*. Budapest: akadémiai kiadó, 1990.

Neubach A. and Cohen A.D.. Processing Strategies and Problems Encountered in the use of Dictionaries, *Dictionaries. Journal of the Dictionary Society of North America,* 10, 1-20, 1988.

Nuccorini S.. On Dictionary Misuse, *Proceedings EURALEX'94* (Amsterdam), 586-597. 1994.

Quillian M.R.. *Semantic Memory*, in M. Minsky ed., 227-270, Semantic Information Processing. Cambridge (Mass.): MIT Press, 1968.

Sager J.C.. *Language Engineering and Translation: Consequences of automation.* Benjamins Translation Lib., 1994.

Schreiber G., Wielinga B., Akkermans H., Van de Velde W.. CML: The CommonKADS Conceptual Modelling Language, *8th European Workshop EKAW'94*. 1994.

Schreiber G., Wielinga B., Breuker J. (eds.). *KADS: A Principled Approach to Knowledge-Based System Development,* Knowledge-Based Systems, vol. 11. London: Academic Press, 1993.

Starren P. and Thelen M.. General dictionaries and students of translation: A report on the use of dictionaries in the translation process, in T. Magay, J. Zigány eds., 447-458, *Proceedings BudaLEX´88*. Budapest: akadémiai kiadó, 1990.

Steiner E.. Producers — Users — Customers: Towards a Differentiated Evaluation of Research in Machine Translation, *Machine Translation*, 7, 281-284, 1993.

Dictionaries

Le Plus Petit Larousse. Paris: Librairie Larousse, 1980.

Aurrekoetxea G. et al.. *Euskararako Hiztegia.* Bilbo: Adorez, 1986.

Lhande P.. *Dictionnaire Basque-Français.* Paris: Gabriel Beauchesne ed., 1926.

Sarasola, I. *Hauta-lanerako Euskal Hiztegia*. Donostia: KUTXA, 1991.

Appendix 1

List of the primitive functions

1. ***anali***: morphological analysis of a word form.
   Input: a word form
   Output: the lemma and the morphological information of the form
2. ***dis_ent***: choice of a dictionary entry in a given context.
   Input: a list of dictionary entries and a context
   Output: a dictionary entry
3. ***dis_con***: choice of a word sense in a given context.
   Input: a list of word senses and a context
   Output: a word sense
4. ***conlist***: list of the possible senses that could be suitable for a word in a given context.
   Input: a word and a context
   Output: a list of the senses associated to the input word
5. ***ddef***: definition request.
   Input: a word sense
   Output: the definition associated to this word sense
6. ***rdef***: reformulation of a definition.
   Input: a word sense, a definition
   Output: another (reformulated) definition for the word sense
7. ***dpro***: request for properties of a concept.
   Input: a word sense
   Output: list of semantic properties of the sense
8. ***dis_def***: choice of a definition in a given context.
   Input: a list of definitions of a word form
   Output: a definition
9. ***ddif***: request for differences for two concepts.
   Input: two word senses
   Output: a list of the differences between the two senses
10. ***drap***: request for relationships between two concepts.
    Input: two word senses
    Output: a list of the relationships between the two senses
11. ***vrap***: verification of relationships between two concepts.
    Input: two word senses and a list of relationships
    Output: true iff the relationships between the two senses are correct
12. ***vdef***: definition verification.
    Input: a word sense and a definition
    Output: true iff the definition is correct for the word sense
13. ***vpro***: verification of properties of a concept.
    Input: a word sense and a list of properties
    Output: true iff the word sense holds the properties
14. ***rths***: thesaurus-like search of concepts.
    Input: a list of constraints
    Output: a list of concepts that satisfy the set of constraints
15. ***exam***: request for examples.
    Input: a word sense
    Output: a list of examples of usage
16. ***trans_lex***: direct lexical translation of a word form.
    Input: a word form
    Output: a translation of the input word form into the target language
17. ***vequiv***: verification of translation equivalents.
    Input: two word senses
    Output: true iff one word sense is the translation of the other
18. ***comp_sem***: semantic compatibility between two word senses according to a given relation.
    Input: two word senses and semantic relation
    Output: true iff the word senses are compatible according to the relation
19. ***synt_pat***: search for syntactic constructions that correspond to a given pattern.
    Input: a syntactic pattern
    Output: a list of syntactic constructions that correspond to the pattern
20. ***colloc***: lexical collocation.
    Input: a word sense and a relation
    Output: a list of word senses that may be used as lexical collocations of the input word
21. ***dis_prop***: choice of a produced lexical form.
    Input: a list of word forms produced in the target language
    Output: a word form
22. ***prod***: lexical form production.
    Input: a lemma and a list of morphological information
    Output: a word form, which is generated from the lemma and the morphological information
23. ***ver-reg***: request for the verb-regime.
    Input: a verb
    Output: information of the regime of the verb
24. ***equiv***: search for potential translation equivalents.
    Input: a word sense in the source language
    Output: a list of equivalent word senses in the target language