DESIGN OF A TRANSLATOR-ORIENTED DICTIONARY: ENHANCEMENT OF A DICTIONARY KNOWLEDGE BASE BY TASK MODELLING

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

Informatika Fakultatea, Euskal Herriko Unibertsitatea
P.K. 649. 20080 DONOSTIA - Basque Country
jiparipx@si.ehu.es

Abstract

A way of assisting translators in the task of translating lexical units is by offering both usable lexical knowledge and knowledge about the process of translation. In our opinion, both types of knowledge could be incorporated into a dictionary system. This idea of the "active dictionary" has led us to design the MultiLingual Dictionary System (MLDS), which is conceived as a computational dictionary-based help-system for human translators.

Keywords: translation aids, dictionaries, multilinguality, lexicography, expert systems.

1 Introduction

Translators are often involved in the task of choosing suitable target lexical units that correspond with those that are in the source text. Such a task may not be easy or even simple. Rather, the behaviour of human translators when choosing lexical translations reflects complex cognitive processes in which intelligence is needed. A way of assisting translators in this task is by offering both usable lexical knowledge and knowledge about the process of selection. In our opinion, both types of knowledge could be incorporated into a dictionary system, which would be an active tool able to deal with the task of lexical choice in its whole dimension, rather than being only a repository, more or less structured, of words and definitions.

This idea of the "active dictionary" (Martin & Al 90; Fillmore & Atkins 94) has led us to design the MultiLingual Dictionary System (MLDS), which is conceived as a computational dictionary-based help-system for human translators. Its aim is to assist human translators in the task of translating lexical units.

During the conception of the system two levels have been distinguished: a translation-oriented level, where MLDS deals with the strategies and tasks carried out by human translators when translating a text, and a domain level, in which lexical information extracted from dictionaries is represented and exploited.

Our starting point in the design process was the observation, elicitation, and modelling of expertise. The conceptual model thus obtained has become executable in MLDS in the way that modern Knowledge Engineering advocates and supports. We have followed the KADS methodology (Schreiber et al. 93) in the design of the system. The KADS design principles allowed to clearly distinguish the levels of knowledge that MLDS manages.

Functional aspects of MLDS are similar to those of active help-systems (Fischer et al. 84). It is expected that MLDS will give more useful and intelligent answers to translators' queries, recognising their goals and anticipating their needs.

2 Elicitation of the real use of the dictionary into a model

From our point of view, translation tools can not be correctly made without the co-operation of human translators (Sager 94).

Therefore, any attempt to incorporate task-dependent behaviour into a dictionary system should begin with a study of the tasks involved and the users' interactions with the dictionary.

Certainly, it would be desirable to have a well-founded theory about these kinds of uses and interactions. We should not forget that the use of dictionaries has been previously investigated from different perspectives and that interesting studies related to this topic have been published. See (Hatherall 84; Hartman 85; Starren & Thelen 90), and references therein. However, a general theory has not been presented to date.

Our practice has been fundamentally empirical. We have not limited our work to a questionnaire-based method to collect information. Rather, the method we have used is based on direct observation and personal interviews. Following we briefly describe both:

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: the characteristics a dictionary should have in order to be useful when translating, the interest about having computerised dictionaries and their main functionality, and so on.

3 Translation-oriented level: the Task-Structure

The model of expertise obtained in the elicitation process has been specified in the task-structure module of MLDS (see figure 1).

The knowledge contained in the task-structure reflects different strategies, tasks and ways of using dictionaries by translators (Agirre et al., 94a).

The description of these tasks has been made in CML (Conceptual Modelling Language), which is one of the languages used in the CommonKADS methodology (Schreiber et al. 94). Tasks can be divided into composite (uppercase in figure 1) and primitive (lowercase) ones.

The composite tasks are described as non-trivial processes to be decomposed into subtasks when they are carried out. For instance, when the *target_word_production* composite task would be executed, it would take the dictionary user from (pre-lexical) meaning to an appropriate word to express this meaning. Four sub-tasks are involved in it: *finding-production-hypothesis, discriminating-production-hypothesis, production-hypothesis-verification* and *from-the-dictionary-entry-to-the-lexical-unit*. The performance of the task reflects a trial-and-error strategy, where different production hypotheses are considered.

Primitive tasks refer to the primitive actions —or cognitive steps— identified as useful for dictionary-users when trying word translations. For example, the *rths (thesaurus-like search of concepts)* task is one of the twenty five primitive tasks defined in the model. This function represents the search for lexical units when the user has an idea but does not know how to formulate it by means of a word. According to our model, *rths* would be preferentially used when verifying the meaning of a source word or when finding production hypothesis (see figure 1), which are its parent-tasks.

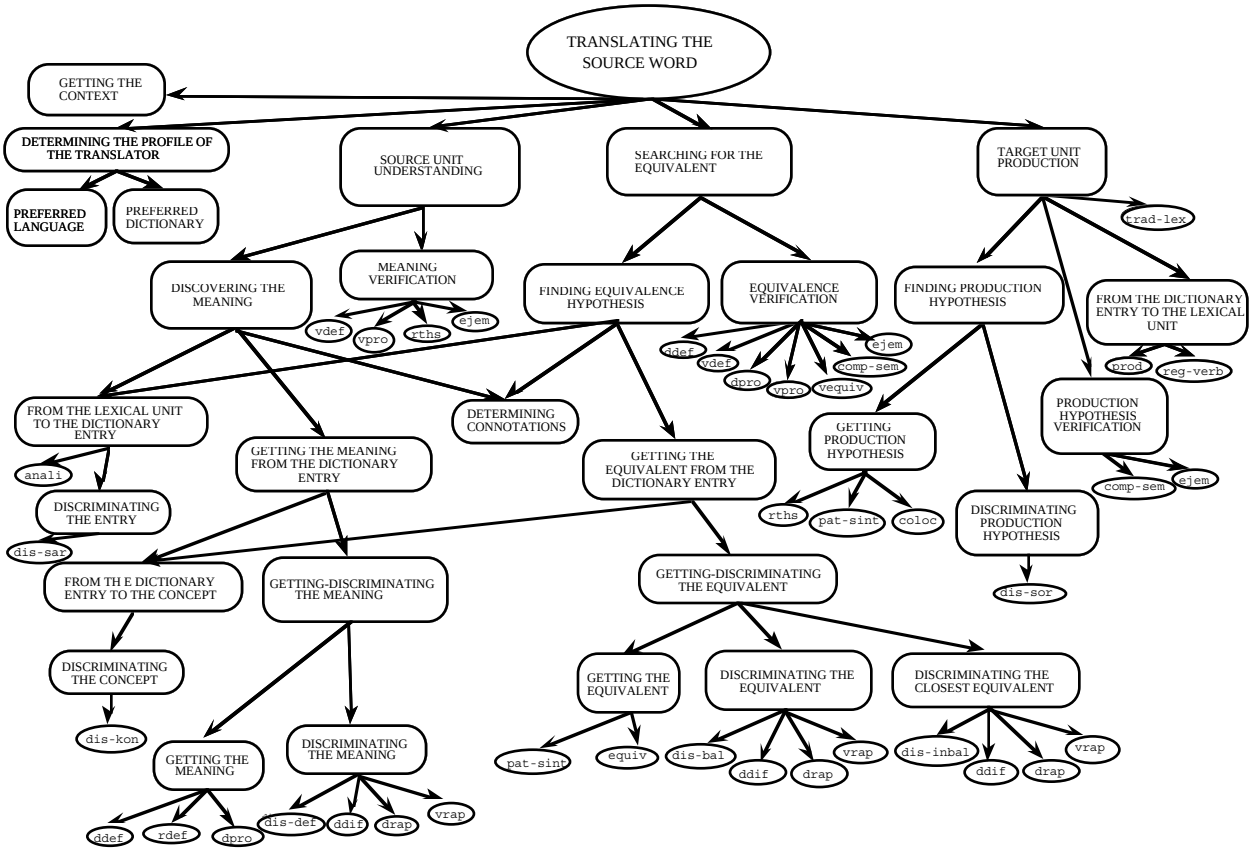

Figure 1: Decomposition-diagram of the tasks involved in the word-level translation process

The conceptual model described in CML becomes executable, as MLDS is able to execute the strategies of the task-structure. Thus, MLDS simulates the behaviour of a human translator when s/he is using the dictionary for solving a lexical problem (Arregi 95).

4 Domain level: the Dictionary Knowledge-Base

The knowledge extracted from conventional dictionaries (Agirre et al. 94b) along with the set of primitive functions (Agirre et al. 93b) constitutes the *domain level* of our system. Note that we are referring to static information (extracted from dictionaries) and functions. These functions could be seen as primitive tasks incorporated by the *active* dictionary. Such a dictionary is a system with capabilities to make inferences on the lexical knowledge.

The knowledge base of MLDS has been built according to the information included in two dictionaries of Basque (HLEH) and French (LPPL), but it is supposed to offer a general framework for different languages. In the present version, the system is composed of four main knowledge bases: SDMOL1 and SDMOL2 contain information concerning the source and object languages respectively, whereas SBL1/2 links concepts from SDMOL1 with concepts included in SDMOL2. S/STRUCTURES includes the representation of the basic objects (attributes, active values, rules).

Both SDMOL1 and SDMOL2 are based on a knowledge representation scheme consisting of three elements, each of them structured as a different knowledge base (Artola & Evrard 92; Artola 93). STRUCT contains meta-knowledge about concepts and relations in D-LPPL (or D-HLEH) and TH-LPPL (or TH-HLEH). D-

LPPL and D-HLEH allow access from the dictionary word level to the corresponding concept level in the DKB. TH-LPPL and TH-HLEH are the representation of dictionaries as semantic networks 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 such as synonymy, taxonomic relations, meronymic relations, specific relations realised by means of meta-linguistic relators, casuals, etc. For more details see (Agirre et al. 94c).

The bilingual module SBL1/2 is composed of two knowledge bases that relate concepts of the source language to concepts of the target language. S/BILINGUAL includes the definition of the classes and attributes needed in the representation of the Bilingual Dictionary. Three different classes have been defined: Source-Unit Class, Target-Unit Class and Bilingual Unit Class. BASQUE/FRENCH contains the information of the Basque/French Bilingual Dictionary. Each bilingual-unit represents an equivalence-link between a source-unit and a target-unit. These links are complemented with information about types of equivalence, equivalence levels, etc.

The implementation proposed for source-units and target-units guarantees the independence of the monolingual environment in relation with the bilingual one.

It is interesting to remark that the connection among different languages permits the exploitation of interlingual relations. These relations besides the intralingual ones allow the inference system to deduce implicit dictionary knowledge. Inferences are activated dynamically when translators interact with MLDS.

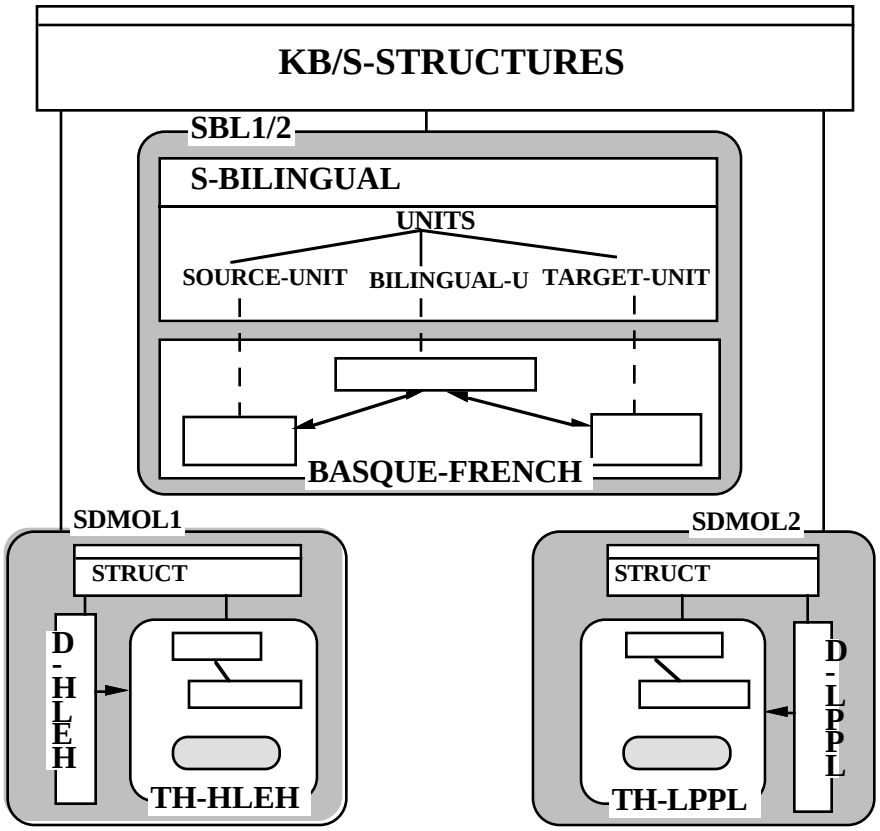

Figure 2: Representation schema of the Dictionary Knowledge Base

5 Functional vision of MLDS

A computerised system with abilities to translate words could be used as an autonomous tool with a limited functionality. Basically, such a tool would translate a given word from a language into other. However, we prefer MLDS to be human assistance-oriented and integrated in a more general framework. Mainly, we are interested in endowing MLDS with the means of interacting with the human translator. This interaction will enrich the performance of the human translator if it preserves some rules:

- The translator must be responsible for the translation of the words.
- MLDS ought to show the real complexity of the task, warning the translator about problems and risks, and proposing different ways of finding, choosing or verifying.
- MLDS must adapt to the state of the translation when interacting.
- MLDS should exploit its deducting and anticipating capabilities.
- MLDS must not be annoying.
- MLDS should combine active and passive help.

According to the mentioned features, three types of help-strategies have been designed.

5.1 Question-answer strategy

This is the strategy followed when the user poses questions i.e., in the passive help mode. The translator requires information to the dictionary system by means of direct queries. There is a correspondence between these queries and the primitive tasks of the conceptual model. Each query activates a primitive function of the inference level in the dictionary knowledge base.

The set of primitive tasks of the system is described in (Agirre et al. 93a). The following is just an example of the use of the RTHS function (thesaurus-like search of concepts):

User.-
RTHS ((and (?X HYPERONYME |consumer I 1|)
(?X AGENT |feu I 1|)),
LPPL, Basque, ?X, ?LC)

The user asks for verbs in Basque that correspond with "to consume with agent fire".

System.- CL=(|izeki I 1|, |kiskali I 1|)

to burn, to blacken.

5.2 The feedback interacting strategy

In this strategy, as in the question-answer one, it is the translator who triggers the basic functions, but in this case the complexity of the function leads the system to ask for help from the user. Usually these kinds of dialogues are due to the need of binding or disambiguating. The user may answer the system in order to improve the performance of the system. However, they are not forced to do so, as any annoying behaviour of the system must be avoided.

5.3 The autonomous behaviour

MLDS decides by itself to help to the user, assuming s/he needs it. This is what is known as active help in the literature. Typically the active help systems assist the user when error situations occur. However, in our case it is not easy to formulate when erroneous dictionary-queries are made. This is why, instead of adopting this approach to active assistance, we have understood autonomous behaviour as a way of giving complementary information whenever possible.

Such an autonomous process is activated when the system detects that a query-sequence has been interrupted without concluding the translation. In this situation MLDS assumes that the translator does not know how to continue and decides to use the knowledge about the translation process that it has.

The complementary information is displayed in a specific window without interfering in the translator's activity. If the translator considers that the complementary information is interesting, s/he uses it, if not, nothing is done.

6 Conclusions

A vision of a human user-oriented dictionary system has been presented. It is expected that this approach will improve the use of lexical information when translating words, incorporating task-oriented knowledge into lexical knowledge.

The MultiLingual Dictionary System has been conceived as a help-system with active abilities. A prototype of the system, which will provide the user with an enriched dictionary assistance, is being developed using the KADS design methodology. In the near future, we are going to test the prototype with human translators.

References:

(Agirre et al. 93a) E. Agirre, X. Arregi, X. Artola, A. Díaz de Ilarraza, F. Evrard, K. Sarasola, *IDHS, MLDS: Towards Dictionary Help Systems for Human Users*, Semantics and Pragmatics of Natural Language: Logical and Computational Aspects, Sara, 167-188, 1993.

(Agirre et al. 93b) E. Agirre, X. Arregi, X. Artola, A. Díaz de Ilarraza, F. Evrard, K. Sarasola, *Sistema diccionarial multilingüe: aproximación funcional*, Boletín SEPLN , 14, 313-333, 1993.

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

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

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

(Arregi 95) X. Arregi, *ANHITZ: Itzulpenean laguntzeko hiztegi-sistema eleanitza / Sistema diccionarial multilingüe de ayuda en la traducción*, Ph.D. Thesis. Univ. of the Basque Country, 1995.

(Artola & Evrard 92) X. Artola, F. Evrard, *Dictionnaire intelligente d'aide à la compréhension*, Actas IV Congreso International EURALEX'90 (Benalmádena), 45-57. Barcelona: Bibliograf, 1992.

(Artola 93) X. Artola, *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.

(Fillmore & Atkins 94) C.J. Fillmore, T.S. Atkins, *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.

(Fischer et al. 84) G. Fischer, A. Lemke, T. Schwab, *Active Help Systems*, in T. Green, M. Tauber, G. van der Veer eds., 116-131, Proceedings of 2nd Conf. on Cognitive Ergonomics. Heidelberg: Springer Verlag, 1984.

(Hartman 85) R.R.K. Hartman, *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.

(Hatherall 84) G. Hatherall, *Studying dictionary use: some findings and proposals*, Proceedings LEXeter'83, 183-189, Lexicographica, Series maior 1, Tübingen; Niemeyer. 1984.

(Martin & Al 90) W. Martin, B.P.F. Al, *User-orientation in dictionaries: 9 propositions*, in T. Magay, J. Zigány eds., 393-399, Proceedings BudaLEX´88. Budapest: akadémiai kiadó, 1990.

(Sager 94) J.C. Sager, *Language Engineering and Translation: Consequences of automation*. Benjamins Translation Library, 1994.

(Schreiber et al. 93) G. Schreiber, B. Wielinga, J. Breuker (eds.), *KADS: A Principled Approach to Knowledge-Based System Development*, Knowledge-Based Systems, vol. 11. London: Academic Press, 1993.

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

(Starren & Thelen 90) P. Starren, M. Thelen, *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.