ANALYSING WORD-LEVEL TRANSLATION ACTIVITY TO DESIGN A COMPUTERISED DICTIONARY HELP SYSTEM

1 Introduction.

Translation, even word-level translation, is not a trivial activity. Human translators are concerned with a wide variety of problems, doubts and questions, and hence, they are compelled to solve complex tasks in order to achieve the goal, i.e. to get correct translations.

Translators have to rely on their own knowledge and experience, being dictionaries the most useful tools. Martin W. and Al B.P.F. (1988) remarked: "*...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*". Unfortunately, conventional dictionaries lack these dynamic features. In this paper a target-oriented computer approach is proposed to improve the use of dictionaries, adapting them to human translation activity.

MultiLingual Dictionary System (MLDS) is conceived as a computational dictionary-based help system for human translators. The functional behaviour of MLDS (Agirre *et al.*, 93) has been defined following a task-based methodology. By studying the use of dictionaries in the translation process a model of the task structure has been designed. This model enables MLDS to give more useful and intelligent answers to translators' queries, recognising their goals and anticipating their needs.

Meaning definitions from two monolingual dictionaries (French and Basque) have been analysed and interpreted (Artola X., 93) in order to constitute the monolingual knowledge bases (KB) of MLDS. These are related by means of bilingual KBs —representing bilingual dictionaries— that establish equivalence links among concepts of the monolingual ones. Each monolingual KB is represented as a semantic network of concepts interrelated by attributes representing lexical-semantic relations such as taxonomy, synonymy, meronymy, and specific relations derived from the lexicographic metalanguage used in definitions.

Following is given an overview of the analysis of dictionary use. Section 3 presents briefly how the word-level translation is structured, some behavioural aspects involved in this task and the basic functions that MLDS supports. The system as a help tool is outlined in section 4. Some conclusions will be presented in the last section.

2 A study on translator needs and dictionary use.

Dictionary use has been investigated from different perspectives and using several methodologies. Interesting studies related to this topic have been presented (Ard J., 1982;

Hatherall G., 1984; Hartman R.R.K., 1985; Atkins B.T. and Knowles F.E., 1990; Starren P. and Thelen M., 1990).

Traditionally three different methods have been used in the analysis of dictionary use: a) free invention that relies only on intuition and speculation, b) questionnaires posed to human users, but as Glyn Hatherall (Hatherall, 84) wonders: *"Are subjects saying here what they do, or what they think they do, or what they think they ought to do, or indeed a mixture of all three?"*, and finally, c) direct observation, currently the most used method.

Our method is based on:

a) Direct observation: given several texts to be translated (in our case French and Basque texts) and several dictionaries (with different characteristics), translator's problems, resolutions, and tasks have been recorded. These protocols can be considered as directed by the observer. The aim is to characterise the task of human translators observing the activity of translating words, expressions, context-dependent phrases and even paragraphs (rarely). Each time the human translator looks up in a dictionary, the unit to be translated, the dictionary used, the dictionary entry and the type of consultation are recorded.

b) Personal interviews with professional translators. These interviews have allowed us to consider 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 functionalities, and so on.

3 Modelling word-level translation.

As a result of the analysis of dictionary use, the word-level translation task has been modelled. It is not, of course, an easy matter to discover how users translate, even dealing with word-for-word translations, yet some interesting insights can be gained.

A distinction between expert translators and not so advanced ones has been revealed. While the former base the translation on the understanding and generation process by using monolingual dictionaries whenever are needed, the latter prefer to use bilingual dictionaries and find the translation equivalents as soon as possible.

It will also be adequate to distinguish between translator and lexicographer roles in our model. Obviously both translators and lexicographers translate lexical units, but differences between their methods are relevant. One of the most important difference is that translators deal with words in context. MLDS combines the points of view of translators and lexicographers.

Modelling the process of word-level translation means structuring the possible answers to the following question: How is a lexical unit included in a text translated using dictionaries? We follow a task-analysis methodology with the following steps (Johnson P. and Johnson H., 1986):

First, it is necessary to identify the translator's goals, subgoals and tasks. Second, the order in which subtasks are carried out must be considered. Third, the different task strategies have to be defined indicating the circumstances under which those strategies are employed. Finally, procedures related with the objects involved in the tasks must be specified in terms of primitive actions. The structure obtained configures the control-knowledge of MLDS.

In our case these primitive actions have already been designed and implemented (Agirre *et al*., 93), and constitute the basis of the functionality of MLDS. These primitive functions are used in the different subtasks involved in the word-level translation, which is the top-level task.

Based on the knowledge acquired from our empirical studies, a model of the translation process is proposed in terms of the subtasks and the primitive functions involved in it. At the top-level three main subtasks have been identified, namely source unit understanding, searching for the equivalent and target unit production. Our experience with translators show us that these subtasks do not occur in a rigid order. At the bottom of the task structure the primitive functions of MLDS are found.

Following the structure of the main subtasks is presented, focusing on the first two steps of the mentioned task-analysis methodology.

Source text understanding

In order to get the meaning of the source text-word, translators use, besides definitions, context information, metaforicity and usage labels. Frequently, connotations are text-range features rather than word-features; however, texts are exclusively composed by words, so there will be words in the text that are marked with connotations. This kind of information extracted from those words is not only relevant for them, but also for all the other words in the text. Consequently, it is necessary to gather —in the source text understanding task— and to use —whenever it is needed during the translation— context-level information.

The figure below presents a typical strategy used by human translators in their activity of understanding a context-dependent source word. Boxes represent subtasks. Arrows mean hierarchical relations among tasks and they can be understood as: all daughter subtasks have to be performed in order to carry out the parent task. Sequential relations among subtasks are not presented in the figure.

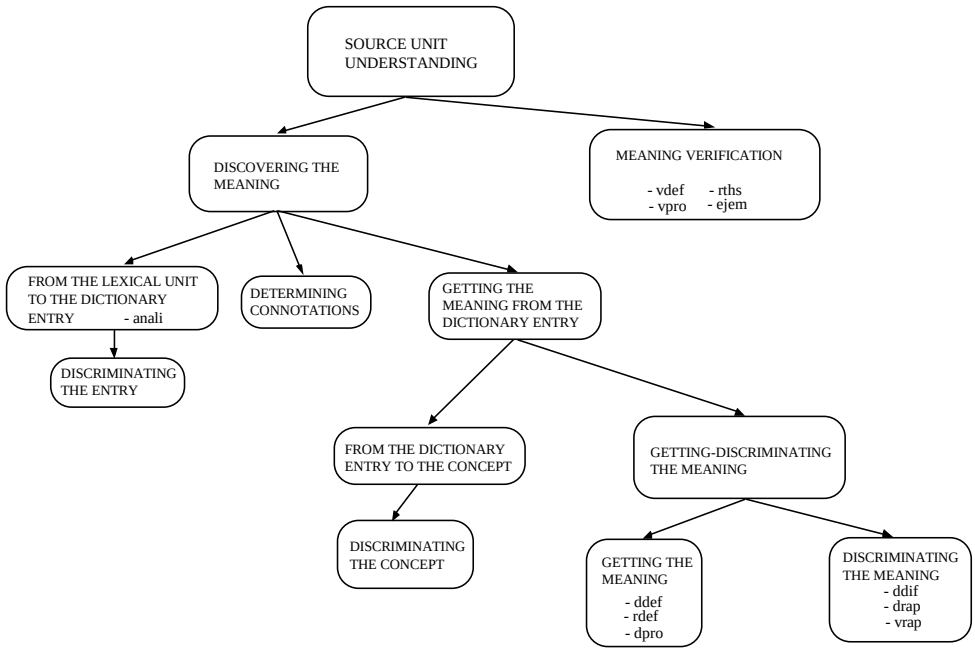

Fig. 1 : Plan associated to the Source Word Understanding task.

Note that the first step consists in getting the dictionary entry from the text-word. Then, the translator tries to choose correct senses and adequate definitions among all possible ones. The translator must solve here difficult problems, and so context information and deep language knowledge have to be used.

When a definition has been chosen this hypothesis is verified before following. The translator is involved in a backtracking process, in the sense that he/she may eventually searches for an alternative definition, sense or even for another entry if he/she is not satisfied with the results obtained.

Function identifiers (in lowercase in the figure) correspond to primitives of MLDS which are performed as part of the subtasks that include them. A brief outline of them is presented here. For a thorough explanation refer to (Agirre *et al*., 93):

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)* in the LPPL French dictionary, with *inherited* as explanatory-level. The result is the textual definition plus other information that, not being explicit in the dictionary (in italics in the example), is deduced by MLDS.

```
Translator.-    DDEF (|guêpe I 1|, inherited, LPPL, French, ?D)
MLDS.-          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 of properties of a concept (DPRO), verification of properties of

a concept (VPRO), request of differences for two concepts (DDIF), request of relationships between two concepts (DRAP), verification of relationships between two concepts (VRAP), thesaurus-like search of concepts (RTHS), morphological analysis of a word form (ANALI) and request of examples (EJEM).

Searching for the equivalent

Getting and discriminating equivalents is one of the most difficult and controversial work translators have to do. Figure 2 reflects this activity.

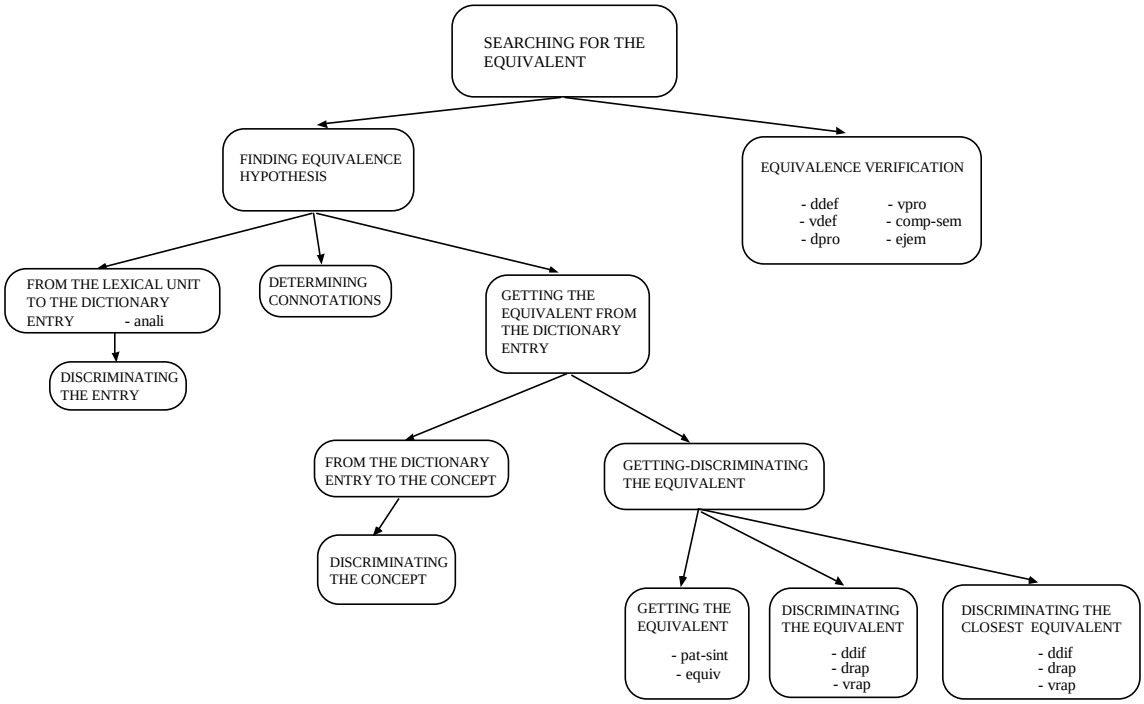

Fig. 2: Plan associated to the Searching for the Equivalent task.

Novice translators tend to get the equivalent as soon as possible, therefore, they may not have understood the meaning of the source word before trying to get an equivalent. Because of this, the tasks related to the choice of correct entries and senses appear in the searching for the equivalent task as well.

The absence of lexical equivalents is a well known problem often discussed by lexicographers and translators. In MLDS two kind of lexical gaps are considered: (a) when there is no single word in the target language to express the source concept, which can be solved by means of translational explanation or *phrasal concept equivalents* in terms of our domain knowledge representation, and (b) when the source concept does not appear as an entry in the bilingual dictionaries; in this case, our alternative relies on finding the closest equivalent and using set operators as ? and • to express that the target concept is respectively *more general* or *more specific* than the source concept.

Once an equivalent has been found, the translator attempts to verify the equivalence-correctness by verifying its concordance with context.

MLDS offers a set of basic functions in order to accomplish the task of searching for a suitable equivalent: search for potential translation equivalents (EQUIV), search for syntactic constructions that correspond to a given pattern (PAT_SINT), semantic compatibility (COMP-SEM), …

Among all these functions, search for potential translation equivalents is particularly relevant. In the first two examples below |*pattar I 1*| and |*txakolin I 1*| are not in the bilingual dictionary, and therefore the system gives the closest concept from the monolingual dictionary and indicates whether it is more or less specific. In the last example there is no single word to say *abere* (domestic animal) in French, and a phrasal concept is returned.

```
Translator.- EQUIV (|pattar I 1|, French, common, ?LP)
MLDS.-         LP =  (•, |eau-de-vie I 1|)
Translator.- EQUIV (|txakolin I 1|, French, common, ?LP)
MLDS.-         LP =  (?, |vin I 1|)
Translator.- EQUIV (|abere I 1|, French, common, ?LP)
MLDS.-         LP =  (|animal I 1#n|)
where |animal I 1#n| represents "domestic animal".
```

Target unit production

Before producing the target word the meaning of the source unit has already been discovered and, in some cases, an equivalent has been chosen. Expert translators often produce the translation result without looking up bilingual dictionaries, but only understanding the source unit and formulating it straight into the target language. Conventional dictionaries do not help enough in the latter, namely in discovering a word from an idea. However, this way of translating is not common for less advanced translators, who prefer to produce directly the target unit on the basis of a chosen equivalent from a bilingual dictionary. Novice and expert have to verify hypothesis and transform an equivalent concept into a context-dependent word.

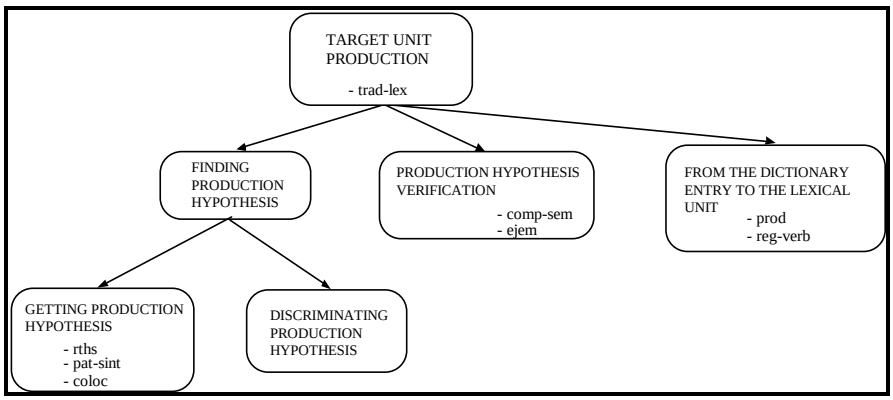

Fig. 3: Plan associated to the Target Word Production task.

It is worthwhile to underline several basic functions involved in this task: thesaurus-like search of concepts (RTHS), lexical collocation (COLOC), lexical form production (PROD), verb-regime (REG-VERB), word translation (TRAD-LEX), …

For instance, RTHS takes as input a restriction-expression, a dictionary and a language, and returns the list of concepts that meet the restrictions stated.

```
Translator.-    RTHS((and   (?X   HYPERONYME   |consumer I 1|)
                            (?X   AGENT        |feu I 1|)),
                            LPPL, French, ?X, ?LC)
                The user asks for verbs in French for to consume with
                agent fire
MLDS.-          LC=(|brûler I 1|, |embraser I 1| )
                to burn, to blacken.
```

4 MLDS as a translation help system.

MLDS is conceived as an interactive help system that tries to answer satisfactorily the questions posed by translators. These queries are formulated in terms of a set of primitive functions offered by the system When MLDS is not able to find a correct answer, it establishes a dialogue with the user, who guides the system to the answer.

The translation process model represents different ways to get equivalent words. The application of this model has to be flexible enough so that the user can choose his/her own way to solve any translation task. MLDS does not aim to lead the user, just to help him/her answering the posed questions and anticipating user expectations.

During the translation, user-system interchanges are memorised in order to set up a translator profile. Aspects like preferred language and dictionary are fundamental to select default strategies.

5 Conclusions

MLDS, MultiLingual Dictionary System, has been presented as a human translator-oriented tool. The domain knowledge of MLDS has been acquired by analysing meaning definitions of several dictionaries, so that several knowledge accessing capabilities are provided.

In order to establish the basis of the control-knowledge of MLDS, i.e. to determine how different subtasks have to be performed to produce correct translations, an study of the lexical translation process has been carried out. The methodology followed in this study is based on direct observation of the tasks of human translators and on personal interviews to the experts trying to characterise the typical use of dictionaries in translation.

A prototype of a French Monolingual environment has been implemented on a Symbolics Lisp machine using KEE (Knowledge Engineering Environment). The knowledge

base of this monolingual system prototype built is composed by 6003 concepts, obtained from the definitions of a French dictionary. We are now working on the multilingual knowledge representation, integrating two monolingual dictionaries (French and Basque) and a bilingual one in the KBs of MLDS, and analysing the different relations needed for the implementation of the described functional behaviour. A set of primitive functions has already been implemented.

Acknowledgements

We would like to thank to the translators Edorta Agirre, Olatz Arregi, Itziar Baragaña, Bittori Fernández, Antxon Gurrutxaga, Xabier Mendiguren and Miriam Urkia, who have contributed to this project.

This work has been sponsored by the *Gipuzkoako Foru-Aldundia*. Eneko Agirre was enjoying a grant from the Basque Government during the present work.

References

Agirre E., Arregi X., Artola X., Díaz de Ilarraza A., Evrard F., Sarasola K.. Sistema diccionarial multilingüe: aproximación funcional, *SEPLN Congress*, Santiago, Spain. 1993.

Ard J.. The use of bilingual dictionaries by ESL students while writing, ITL Review of Applied Linguistics, 58, 1-27.1982.

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

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

Atkins B.T. and 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.

Boguraev B., Briscoe T.. eds. *Computational Lexicography for Natural Language Processing.* New York:Longman**,** 1989.

Byrd R.J., Calzolari N., Chodorow M.S., Klavans J.L., Neff M.S., Rizk O.A.. Tools and Methods for Computational Lexicography, *Computational Linguistics* 13, 3-4, 219-240. 1987.

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.

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

Heylen D., Maxwell K.G., Armstrong-Warwich S.. Collocations, Dictionaries and MT, *AAAI93 Spring Symposium Series "Building Lexicons for Machine Translation"* (Standford), 1993.

Johnson P. and Johnson H.. Knowledge analysis of tasks: task analysis and specification for human-computer systems, in A.Downton ed., 119-144, *Engineering the human-computer interface*. London: McGraw-Hill, 1991.

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.

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.

Dictionaries

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

Aurrekoetxea G., Badiola R., Baraiazarra L., Goikoetxea J.L., Solabarrieta T., Uribarren P. *Euskararako Hiztegia.* Bilbo: Adorez, 1986.

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