

# Rule-based Acquisition and Maintenance of Lexical and Semantic Knowledge \*

Donna M. Gates and Peter Shell  
Internet: dm@cmu.edu, pshell@cmu.edu

Center for Machine Translation  
Carnegie Mellon University  
5000 Forbes Avenue  
Pittsburgh, PA 15213  
U.S.A.

## Abstract

The lexicons for Knowledge-Based Machine Translation systems require knowledge intensive morphological, syntactic and semantic information. This information is often used in different ways and usually formatted for a specific NLP system. This tends to make both the acquisition and maintenance of lexical databases cumbersome, inefficient and error-prone. In order to solve these problems, we have developed a program called COOL which automates the acquisition and maintenance processes and allows us to standardize and centralize the databases. This system is currently being used in the ESTRATO machine translation project at the Center for Machine Translation.

## 1 Introduction

In this paper, we describe a fully-implemented rule-based system for the semi-automatic acquisition and maintenance of lexical and semantic knowledge in a knowledge-based machine translation system. This rule-based system is called COOL: *Creator of ontologies and Lexicons*. COOL can create and update various lexical and semantic *knowledge sources* for different NLP modules.

COOL is a working system that was developed for ESTRATO (EScuela de TRAductores de TOledo), a joint project of the Center for Machine Translation at CMU and Union Electrica Fenosa, an electric utility company in Madrid, Spain. ESTRATO is a system

for translating Spanish to English in a restricted domain with controlled input. ESTRATO consists of several modules from the KANT MT system [Mitamura *et al.*, 1991] as well as morphological analysis and phrasal recognition modules and the TWS authoring environment [Nirenburg *et al.*, 1992].

As shown in figure 1, every ESTRATO run-time module uses a different lexical or semantic knowledge source, which differs in content as well as format. The knowledge contained in these modules overlaps. We needed to coordinate and maintain these knowledge sources in a robust and efficient way. Furthermore, the lexical information needed by the translator is initially acquired by people ("editors") who are neither linguists nor domain experts. This lexical information is kept in *lexical feature files*.<sup>1</sup> We needed a way to convert these lexical feature files into forms which could be used by the run-time modules of ESTRATO.

Our solution is to maintain a centralized *lexical and semantic frame database*, and to use COOL to help us acquire this database by converting the initial feature files created by the human editors. The lexical and semantic knowledge sources needed by the run-time translator are then automatically generated and maintained by COOL. Two subsystems perform these tasks: ACQUISITION-COOL (A-COOL) and MAINTENANCE-COOL (M-COOL). A-COOL produces the central lexical and template semantic databases from the initial lexical feature files, which the linguist and domain expert can then modify. M-COOL goes beyond simple acquisition of lexical information. It automatically generates efficient run-time versions of the lexical and semantic knowledge from the central repository of lexical and semantic databases initially

\*This project was funded by Union Electrica Fenosa, Madrid, Spain.

<sup>1</sup>We do not describe the lexical acquisition program for acquiring the initial lexical feature files.

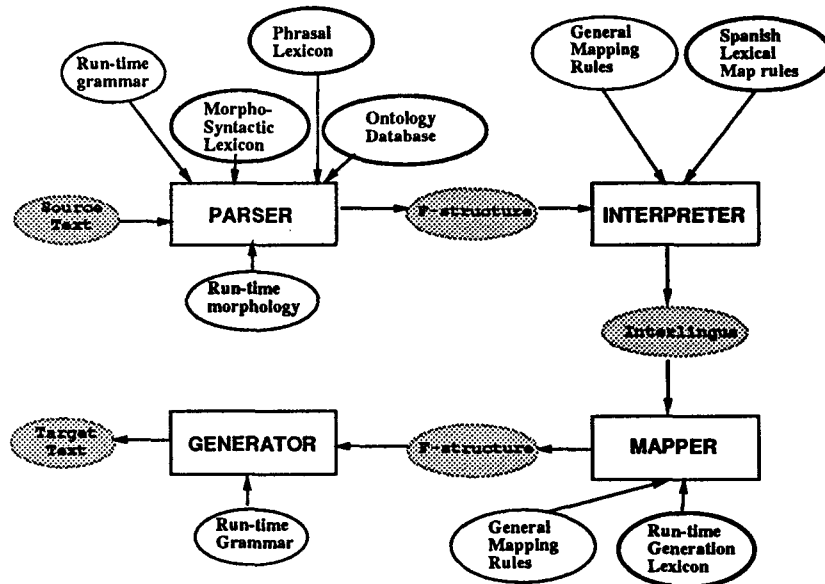


Figure 1: Knowledge-based translator and run-time knowledge. Thick ovals represent knowledge generated by COOL.

created by A-COOL and maintained by experts. The relationship between A-COOL, M-COOL and the three different types of information is depicted in figure 2.

COOL maintains consistency in the knowledge sources and makes it easy to add lexical databases for new modules. By keeping a single source for all lexical information for a given language, COOL allows us to robustly maintain knowledge and eliminate redundancy, by using the power of a frame-based rule language.

First we describe the acquisition and maintenance problems in more detail, and then describe the A-COOL and M-COOL tools which we developed to solve these problems. We also look at related efforts, and mention some ideas for future work.

## 2 The Knowledge Acquisition and Maintenance Problems

At the Center for Machine Translation, we use Lexical Functional Grammar (LFG) [Kaplan Bresnan, 1982] as a basis for our syntactic grammars as well as our linking rules [Levin, 1987] for mapping syntactic functions to and from semantic roles. The latter we refer to as "mapping rules". These mapping rules are used in conjunction with a domain model to build or generate from the interlingua text representations (ILT). The use of ILT is characteristic of the CMT approach to Knowledge-Based Machine Translation [Goodman, 1991; Mitamura *et al.*, 1991; Frederking *et al.*, 1992].

Given the emphasis placed on the lexicon in LFG in both syntax and semantics and the extensive domain knowledge required for our translation system, we place a great deal of importance on the lexicon and finding easy methods to acquire, maintain, view, store and reuse the lexical information. COOL is a tool we are developing and using on the ESTRATO

project for accomplishing these tasks.

The knowledge acquisition and maintenance tasks can be rather cumbersome. Acquiring 1000's of new semantic concepts and placing them into the top-level semantic hierarchy by hand is tedious and error-prone. This also applies to adding English and Spanish words. Once the run-time knowledge sources for the various NLP modules have been acquired, maintaining consistency among the lexical and semantic files (phrasal-noun list, glossary, morpho-syntactic lexicons, word-to-concept mappings and the semantic concepts) is difficult. The NLP modules require different lexical and semantic knowledge with varying formats. All modules share some information which must be kept consistent, such as the part of speech and the word-sense. The concept name must be the same for the run-time semantic knowledge source, the Spanish run-time lexical knowledge source, and the English run-time lexical knowledge source. Both acquiring the knowledge and maintaining consistency in the knowledge are prone to human error.

One of the requirements of ESTRATO is that a non-linguist lexicographer be able to acquire and maintain lexical information as much as possible. A-COOL allows the semi-automatic creation of NLP lexical knowledge from lexicographic information supplied by a non-linguist.

At present, linguists must do some of the lexical acquisition work such as providing semantic class information and some specialized syntactic information for closed-class items, adjectives and verbs. When there is not always a one-to-one lexical mapping from a Spanish and English word to the same concept [Talmy, 1972; Talmy, 1985], the lexical entries can only be produced semi-automatically. Linguists must also provide collocational information in

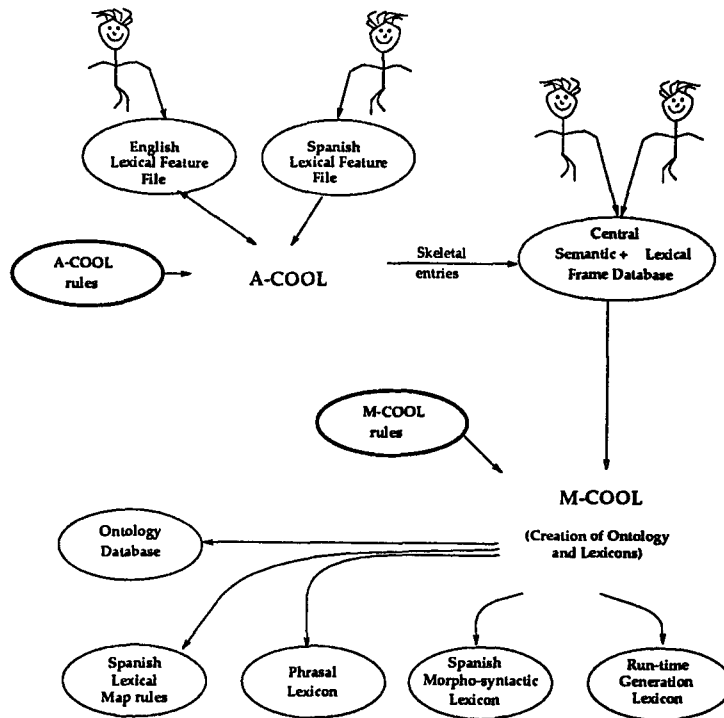


Figure 2: Relationship between A-COOL, M-COOL and lexical and semantic information.

the lexicon relevant to lexical selection [Mel'čuk *et al.*, 1984].

### 3 Automated Knowledge Acquisition

A-COOL automates the acquisition of lexical and semantic knowledge in ESTRATO. For each entry in a Spanish lexical feature file, A-COOL creates: a new semantic concept frame for the central semantic database, a Spanish lexical frame for the Spanish central lexical database and a skeletal entry for the English lexical feature file. Once the entry from the English lexical feature file has been filled out by the editor, A-COOL will also create a lexical frame for the English central lexical database. The word-to-concept mappings for the Spanish and English words are automatically created by A-COOL in order to ensure consistency. A-COOL accomplishes all of this by means of easily modified if-then rules.

When A-COOL creates a new concept, it automatically makes a link to a more general semantic class. The top-level hierarchy we are currently using was created at Carnegie Mellon University [Carlson and Nirenburg, 1990]. The insertion of semantic concepts into a hierarchy is not dependent on the specific top-level. The rules specify the linking of the new concepts in the semantic hierarchy based on features (such as ACTION for verbs and ANIMACY for nouns) in the lexical feature files. These rules can be modified easily for adding concepts to a different top-level.

What follows is a description of the A-COOL process using the entry for the Spanish verb "funcionar" ("to work"). The verb feature ACTION in the lexical acquisition phase is designed such that the user is

```

("FUNCIONAR"
 (cat v) (trans intrans)
 (action physical)
 (eng "function")
 (stem-change no) (comp-type 0)
 ...)

("WORK"
 (cat v) (action physical)
 (span "funcionar")
 (comp-type 0) (trans intrans)
 ...)

```

Figure 3: Sample input verbs to A-COOL

prompted for a response to a question about the type of action the verb represents (if any at all). With this information, A-COOL can produce the preliminary value of IS-A for a semantic frame when it creates the semantic frame from the verb entry.

The "if" or "LHS" (left-hand-side) part of the A-COOL rules specifies properties of lexical features which must be true for the rule to apply. If the rule does apply, the "then" or "RHS" (right-hand-side) specifies which slots of the central database frame to create.

For example, figure 3 shows entries for the Spanish verb "funcionar" and its corresponding English verb "work" from the lexical feature files.

In order to convert these entries into central database frames, the following rules apply. rule1 inserts the default information that the value of the CLASS feature for "funcionar" is AGENT, because the

```
(SPANISH-RULE rule1
  LHS (trans intrans)
      (reflexive unknown)
  RHS
      (class agent)
      (is-a +w-spanish-intrans-verb)
      (trans intrans))
```

Figure 4: A-COOL rule to convert Spanish words

```
(ENGLISH-RULE rule2
  LHS (trans unknown)
      (cat v)
  RHS
      (class agent-theme)
      (trans trans))
```

Figure 5: A-COOL rule to convert English words

reflexive value is unknown (see figure 4). It also inserts the word into the lexical hierarchy under +W-SPANISH-INTRANS-VERB and copies the TRANS information to the new frame.

Similarly, rule2 (see figure 5) helps to convert "work" by guessing at the value of the TRANS slot and setting the CLASS to AGENT-THEME.

Finally, rule3 (see figure 6) helps to generate the template semantic frame corresponding to the meaning of "funcionar" and "work" by placing the frame under PHYSICAL-EVENT in the semantic IS-A hierarchy.

A-COOL works by using the following algorithm:

1. Read in the (Spanish or English) lexical feature file.
2. For each lexical item, generate a frame by applying all relevant rules to that lexical item.
3. Write that frame to the central frame file.

With "funcionar" and "work" as the input lexical items, the rules generate the central frames shown in figure 7.

## 4 Automated Knowledge Maintenance

### 4.1 Introduction

M-COOL allows the linguist to keep just one source for Spanish lexical information and one source for English lexical information (the central lexical frame

```
(SEMANTIC-RULE rule3
  LHS (action physical)
  RHS (is-a physical-event))
```

Figure 6: A-COOL rule to place a semantic frame in the IS-A hierarchy

```
(MAKE-FRAME +W-SP-FUNCIONAR-V-2
  (COMP-TYPE no)
  (CAT V)
  (STEM-CHANGE no)
  (TRANS intrans)
  (IS-A +w-spanish-intrans-verb)
  (CLASS agent)
  (HEAD *work-funcionar)
  (ROOT "funcionar"))
```

```
(MAKE-FRAME +W-EN-WORK-V-1
  (ROOT "work")
  (HEAD *work-funcionar)
  (COMP-TYPE no)
  (CLASS agent)
  (IS-A +w-english-verb)
  (TRANS intrans)
  (CAT V))
```

```
(MAKE-FRAME *WORK-FUNCIONAR
  (IS-A device-event)
  (GOAL *none*)
  (LOCATION building place ...)
  (INSIDE-OF *cabinet-armario ...))
```

Figure 7: Lexical and Semantic frame entries generated by A-COOL and used as input to M-COOL.

databases). Thus, the lexical information is not spread out over several files and can be modified easily. Each language's lexicon can also be organized hierarchically.

Using a set of if-then rules, M-COOL automatically produces the necessary run-time lexical and semantic knowledge sources for the various NLP modules. These rules specify which features are needed for the different modules. The rules also create some lexical knowledge that can be extracted from the lexical and semantic hierarchies. This information need not be specified in the lexical entries.<sup>2</sup>

Since the various run-time lexical and semantic knowledge sources now come from common central databases, consistency is maintained and human error is minimized. Both the semantic knowledge and the lexical knowledge are stored in a standard frame-based format. This allows the linguist and domain-expert to view or modify the knowledge with a frame-based editor.

The rest of this section describes the M-COOL program, the lexical and semantic frames used by M-COOL, and then gives an annotated example to illustrate how M-COOL works.

### 4.2 Program Description

In order to make the knowledge maintenance cycle faster, M-COOL can also work incrementally as well as in batch mode. If the linguist only modifies or

<sup>2</sup>E.g., the linking of syntactic arguments to semantic roles.

adds a small number of lexical or semantic items, the incremental version of M-COOL will only update the run-time knowledge sources which are affected by the changes, instead of re-generating all of the run-time knowledge sources. This saves considerable time over the non-incremental method.

M-COOL works by first determining which run-time knowledge sources need to be updated. For each such knowledge source, it then applies all rules which are relevant to that knowledge source. Each rule is associated with a specific knowledge source.

To extend M-COOL to generate the run-time knowledge source for a new NLP module, two steps are taken:

1. Define the properties of the new knowledge source in the *file-type table*.
2. Write a new set of rules for generating the entries which comprise the new knowledge source. These rules specify the lexical features to be used for the entry as well as the format of the entry.

The *file-type table* simply tells M-COOL whether the given knowledge source is lexical or semantic, and whether it is for generation or analysis. It also supplies miscellaneous information such as the name of the file where the run-time entries are kept and whether it can be compiled using the LISP compile command. For example, our *Spanish-lexical-analysis* file-type is defined with this entry:

```

DATABASE Spanish-lexical-analysis
"Spanish/Mappings/lex-map.lisp"
:lexical :analysis

```

The rule language used by M-COOL is called FRULEKIT [Shell and Carbonell, 1986]. FRULEKIT is an efficient CommonLisp pattern matcher with several extensions over OPS-5. The most relevant extension is that it allows rules to flexibly match against and modify frames in a hierarchy. Having such a frame-based rule language makes it easy for us to write rules to update the ESTRATO runtime knowledge sources.

#### 4.3 Lexical and Semantic Frame Description

Let us briefly discuss the lexical and semantic database files which are the input to M-COOL. The lexical frames are the repository of all lexical knowledge for the ESTRATO system. These frames contain structural, grammatical and some semantic encoding information for words or phrases. They can be easily extended to include other lexical information (e.g., definitions or synonyms) for display to a human translator. For the purposes of ESTRATO, each lexical entry contains a part of speech (CAT), a lexical mapping rule (HEAD or SEM-MAP), a root form (ROOT) and a link (IS-A) to its location in the lexical hierarchy. Nouns (CAT N) contain agreement

```

(MAKE-FRAME +W-EN-GO-OFF-V-1
 (ROOT "go")
 (HEAD *work-funcionar)
 (PATTERN (agent
           (is-a *alarm-alarma)))
 (SEM-DOMAIN "mech/tech")
 (COMP-TYPE no)
 (CLASS agent)
 (IS-A +w-english-verb)
 (TRANS intrans)
 (IRREGULARS (past "went")
             (pastpart "gone"))
 (PARTICLE off)
 (CAT V))

```

Figure 8: *Alternative English lexical entry for \*WORK-FUNCIONAR*

(GENDER and NUMBER) count/mass (COUNT) and a trinary distinction of ANIMACY (human, animal, non-living). Morphological information for Spanish is represented in the feature STEM-CHANGE and for both Spanish and English in the features ALLO-FLAG and IRREGULARS. Verbs and adjectives contain features for subcategorization (TRANS, COMP-TYPE) and features for syntactic-semantic argument linking (CLASS, MAPPINGS). CLASS here refers to the type of linking rules a verb or adjective [Levin and Rappaport, 1987] will use for its syntactic arguments (SUBJ, OBJ, OBJ2, XCOMP, and COMP [Kaplan Bresnan, 1982]). Semantic knowledge about the world is stored in a domain model organized in an is-a hierarchy using frames that correspond to the various events (PHYSICAL-EVENT \*ASSEMBLE-MONTAR) and objects (PHYSICAL-OBJECT \*TRANSFORMER-TRANSFORMADOR), relations (AGENT, THEME)<sup>3</sup> between these objects and events and properties (COLOR, SHAPE) in the specific domain [Carlson and Nirenburg, 1990]. The name of each lexical frame represents a single word sense [Meyer *et al.*, 1992]. Examples of lexical frames are shown in figure 7.

Each frame specifies a link to a parent in the lexical hierarchy or the domain model hierarchy (IS-A). This allows lexical entries to be arranged into classes which require similar "mapping rules" [Mitamura, 1989].

Each semantic knowledge database frame in the domain model also specifies the roles which a given concept may have as well as specific restrictions on the fillers of those roles. An example of a semantic frame was shown in figure 7. The information in the databases is used in different forms and combinations depending on the NLP component's needs.

Figure 8 shows a frame which is an alternative English lexical entry for the concept \*WORK-FUNCIONAR.

<sup>3</sup>We make no theoretical claims about the definition of the roles agent and theme [Guerssel *et al.*, 1985; Jackendoff, 1983].

```
(MRULE lex-analysis-Spanish-verb
:LHS
  (=!+w-sp-Spanish-verb
   :head =head
   :root =root
   :class =class
   :sem-map =sem)
(current-file
 :value Spanish-lexical-analysis)
:RHS
(cool-output
 '(:root (gen-frame-name =verb)
  :cat V
  :head =head
  :class =class
  :sem =sem)))
```

Figure 9: M-COOL rule for generating run-time lexical mapping data.

```
(:ROOT "+W-SP-FUNCIÓNAR-V-2"
:CAT V
:HEAD *WORK-FUNCIÓNAR
:CLASS AGENT)
```

Figure 10: Lexical-map entry generated by M-COOL.

The value of the PATTERN slot in this frame (AGENT (IS-A \*ALARM-ALARMA)) is used so that when the AGENT role is filled with an "alarm", the English word selected for generation is "go off" rather than "work".

#### 4.4 Example

Now we will illustrate how M-COOL rules automatically generate various types of run-time knowledge from the frames shown in figure 7. Figure 9 shows a rule for generating lexical mapping information. This rule applies to the lexical frame +W-SP-FUNCIÓNAR-V-2 in order to generate the run-time lexical analysis mapping data depicted in figure 10.

Next we have a rule for generating the run-time Ontology database, which we call "framettes" (figure 11). This rule applies to the semantic frame \*WORK-FUNCIÓNAR (shown in figure 7) to generate the framette as shown in figure 12.

The two previous rules were fairly simple, but M-COOL can perform much more complex computations. For example, in order to generate efficient run-time knowledge which allows the translator to map from interlingua into English feature-structures, M-COOL must find, for each semantic frame, every English lexical frame which corresponds to it. It then combines this correspondence information into a single LISP function which will efficiently perform the mapping at run-time. One of the M-COOL rules responsible for constructing this knowledge is shown in figure 13. In this example, it applies to the semantic frame \*WORK-FUNCIÓNAR. It finds two lex-

```
(MRULE events-onto-rule
:LHS
  (=!event (LABEL =event))
(current-file :value event-framettes)
:RHS
(cool-output
 '(',(cool-frame-name =event)
  (is-a (class-of =event))
  ,(gen-framette-slots =event))))
```

Figure 11: M-COOL rule for generating run-time event framette data.

```
(*WORK-FUNCIÓNAR
 (IS-A DEVICE-EVENT)
 (INSIDE-OF *CABINET-ARMARIO ...)
 (LOCATION BUILDING PLACE ...)
 (GOAL *NONE*))
```

Figure 12: Event-framette generated by M-COOL.

ical frames which correspond to each other: +W-SP-FUNCIÓNAR-V-2 AND +W-EN-WORK-V-1 (see figure 7). The LISP function generated by this rule is shown in figure 14.

## 5 Related Work

Most of the effort in developing software tools for NLP has focused on user interfaces and acquisition of lexical databases from text corpora, but there are very few rule-based systems for knowledge maintenance. [Pin-Ngern *et al.*, 1989] go beyond corpus analysis by augmenting the lexical databases with knowledge supplied by human editors. The Word Manager [Domenig, 1988] is a system for both acquisition and maintenance of morphological knowledge, but its main strength is its user-interface. LUKE [Knight, 1991] is an interactive system which uses several heuristics exploiting the relationship between linguistic and world knowledge to partially automate the acquisition process.

More effort has gone into the acquisition and maintenance of knowledge for expert-systems.<sup>4</sup> The focus of such efforts is to acquire smaller amounts of problem-solving knowledge, which is more complex than the semantic and lexical knowledge used in ESTRATO.

## 6 Future Work

We intend to extend COOL in three directions: by supporting the acquisition and maintenance of lexical and semantic information for new languages, by adding rules for completely automating the acquisition of semantic classes and lexical argument alternations [Bresnan, 1982; Perlmutter, 1983], and by

<sup>4</sup>For example, [Michalski, 1989] contains several articles on these efforts.

```

(MRULE gen-lex-code-English-verb
:LHS
  (need-lex-info (LABEL =need-info)
    :lex-entry =word
    (CHECK (isa-p (pa-class-of =word)
      '+w-EN-English-verb)))
  (have-lex-info (LABEL =have-info))
:RHS
  ...
  (push (list passive-complete-pattern
    pass-syn-entry
    map-code-pass)
    (have-lex-info-glex-entries
      =have-info))
  (push (list complete-pattern
    syn-entry map-code)
    (have-lex-info-glex-entries
      =have-info)))

```

Figure 13: M-COOL rule for generating a run-time English generation mapping function.

```

(DEFUN ENG-LUTHOR-*WORK-FUNCIORAR (ILT)
(COND
  ((IS-A-P-SLOT 'AGENT '*ALARM-ALARMA)
  (LIST '(SYN ((CAT V) (PARTICLE OFF)
    ... (TRANS INTRANS)
    (IRREGULARS
      ((PAST "went")
      (PASTPART "gone"))))
    (ROOT GO)))
    *ENGLISH-AGENT-VERB-MAPPINGS*))
(T (LIST
  '(SYN ((CAT V) ...
    (TRANS INTRANS)
    (ROOT WORK)))
    *ENGLISH-AGENT-VERB-MAPPINGS*))))

```

Figure 14: Part of an english lexical mapping function generated by M-COOL.

improving the functionality of the underlying system itself. Because it is easy to extend M-COOL to generate run-time knowledge sources for new modules, we plan to add, for example: English-analysis lexical tables, Spanish-generation lexical tables, and lexical tables for an external machine-translation system.

We also have plans for integrating the various acquisition and maintenance tools we use in the ESTRATO system (which include A-COOL and M-COOL) into a single incremental lexical acquisition and maintenance program with a user-friendly interface for both experts and non-experts. The interface will prompt the non-expert for information about a word without the user needing to know linguistics. For example, determining the countability of a noun can be done by prompting the user with examples of the word being used in a countable context and non-countable context. This will allow non-experts to add most of the lexical and semantic knowledge. Currently the process of adding or modifying database entries and running A-COOL and M-COOL requires the user to understand both the internal representation of the lexical items and how to run the various programs. An interactive knowledge editor which hides all of the details from the user will make the user's work much more productive and simple.

## 7 Conclusions

Our idea of developing a program to help automate the task of lexical and semantic knowledge acquisition and maintenance has been very fruitful for us. We have realized the following benefits:

- A-COOL and M-COOL make knowledge acquisition and maintenance easier, faster and more robust. By automatically generating template lexical and semantic database entries from the lexical feature files, A-COOL accelerates the acquisition process and eliminates many sources of human error. Similarly, M-COOL eliminates the need to manually update a large number of run-time knowledge sources each time a new lexical entry is added. By using a powerful and efficient frame-matching rule-based system to automatically generate the correct run-time knowledge sources, knowledge-maintenance is faster.
- M-COOL allows us to integrate generation and analysis lexical knowledge. Because M-COOL can generate both analysis and generation lexical knowledge sources from the same central database, this makes it very easy to create Spanish generation and English analysis knowledge sources. This solves the problem of having to maintain separate versions of knowledge for the analysis and generation of the same language.
- It is easy to extend M-COOL to new modules. Although we didn't anticipate it, we were able to use M-COOL to generate and maintain a wide

variety of additional knowledge sources (for example, a custom glossary and a phrasal-lexicon file). M-COOL's design makes this easy.

Given the complexity and size of our machine-translation system, COOL has become an indispensable part of our knowledge acquisition environment.

### Acknowledgements

We would like to thank the members of the ESTRATO project for their help and support: Mildred Galarza, Jose García, Jose Goyeneche, Michael Mauldin and Teresa Rubio. We would also like to thank Lori Levin and Barbara Moore for their comments and suggestions.

### References

- [Mitamura *et al.*, 1991] Teruko Mitamura, Eric H. Nyberg, and Jaime G. Carbonell. An Efficient Interlingua Translation System for Multi-lingual Document Production. In *Proceedings of the Machine Translation Summit III*, Washington D.C., 1991.
- [Nirenburg *et al.*, 1992] S. Nirenburg, P. Shell, A. Cohen, P. Cousseau, D. Grannes, C. McNeilly. Multi-purpose development and operation environments for natural-language applications. In *3rd Conference on Applied Natural Language Processing*, Trento, Italy, 1992.
- [Frederking *et al.*, 1992] R. Frederking, A. Cohen, D. Grannes, P. Cousseau, S. Nirenburg. The Pangloss Mark I MAT System. In *Proceedings of the European Association for Computational Linguistics Conference, Utrecht, The Netherlands, 1993*.
- [Carlson and Nirenburg, 1990] Lynn Carlson and Sergei Nirenburg, World Modeling for NLP. Center for Machine Translation Technical Report 121, Pittsburgh, PA, 1990.
- [Bateman *et al.*, 1990] John A. Bateman, Robert T. Kasper, Johanna D. Moore and Richard Whitney, A General Organization of Knowledge for Natural Language Processing: the Penman Upper Model. March 1990
- [Meyer *et al.*, 1992] Ingrid Meyer, Boyan Onyshkevych, and Lynn Carlson. Lexicographic Principles and Design for Knowledge-Based Machine Translation. Center for Machine Translation Technical Report 118, Pittsburgh, PA, 1990.
- [Pin-Ngern *et al.*, 1989] Pin-Ngern, Strutz and Evens. Lexical Acquisition for Lexical Databases. In *Proceedings of Computing in the 90's Conference*, Kalamazoo, MI, USA, 1989.
- [Domenig, 1988] M. Domenig. Word Manager: a System for the Definition, Access and Maintenance of Lexical Databases. In *Proceedings of COLING Budapest Conference on Computational Linguistics*, Budapest, 1988.
- [Knight, 1991] Kevin Knight. *Integrating knowledge acquisition and language acquisition*. PhD Thesis, Carnegie Mellon University School of Computer Science, Pittsburgh, PA, 1991.
- [Mitamura, 1989] Teruko Mitamura. *The Hierarchical Organization of Predicate Frames for Interpretive Mapping in Natural Language Processing*. PhD Thesis, University of Pittsburgh, Department of Linguistics, Pittsburgh, PA, 1989.
- [Levin, 1987] Lori S. Levin. *Toward a Linking Theory of Relation Changing Rules in LFG*. CSLI Report No. CSLI-87-115, Center for the Study of Language and Information, Stanford, CA, 1987.
- [Guerssel *et al.*, 1985] Mohamed Guerssel, Kenneth Hale, Mary Laughren, Beth Levin, and Josie White Eagle, A Cross-Linguistic Study of Transitivity Alternations. Presented at the parasession on Causatives and Agentivity at the 21st Regional Meeting of the Chicago Linguistic Society, April 1985.
- [Levin and Rappaport, 1987] Beth Levin and Malka Rappaport, The Formation of Adjectival Passives. *em Linguistic Inquiry* Vol.17, No. 4, 623-661, MIT Press, Cambridge, MA, 1986.
- [Michalski, 1989] R. S. Michalski, J. G. Carbonell and T. M. Mitchell, editors. *Machine Learning, An Artificial Intelligence Approach*, Vol. 4. Tioga Press, Palo Alto, CA, 1989.
- [Goodman, 1991] Kenneth Goodman and Sergei Nirenburg, editors. *The KBMT Project: A Case Study in Knowledge-Based Machine Translation*, Morgan Kaufmann Publishers, San Mateo, CA, 1991.
- [Jackendoff, 1983] Ray Jackendoff. *Semantics and Cognition*, MIT Press, Cambridge, MA, 1983.
- [Perlmutter, 1983] David Perlmutter, editor. *Studies in Relation Grammar 1*, The University of Chicago Press, Chicago, IL, 1983.
- [Bresnan, 1982] Joan Bresnan, Polyadicity, Joan Bresnan, editor. *The Mental Representation of Grammatical Relations*, MIT Press, Cambridge, MA, 149-172 1982.
- [Kaplan Bresnan, 1982] Ronald Kaplan and Joan Bresnan, Lexical Functional Grammar: A Formal System for Grammatical Representation, Joan Bresnan, editor *The Mental Representation of Grammatical Relations*, MIT Press, Cambridge, MA: 173-281, 1982.
- [Talmy, 1985] Leonard Talmy, Lexicalization Patterns: Semantic Structure in Lexical Forms, Timothy Shopen, editor. *Language Typology and Syntactic Description*, Vol. 3 Cambridge University Press, Cambridge, MA, 1985.



- [Talmy, 1972] Leonard Talmy. Semantic Structures in English and Atsugewi. PhD Thesis, University of California, Berkely, CA, 1972.
- [Light, 1992] Marc Light. A Computational Theory of Lexical Relatedness. University of Rochester, Computer Science Department, Technical Report 421, Rochester, New York, 1992.
- [Mel'čuk *et al.*, 1984] Igor Mel'čuk, Nadia Arbatchewsky-Jumarie, Leo Elnitsky, Lidija Iordanskaja, and Adèle Lessard . *Dictionnaire explicatif et combinatoire du français contemporain: recherches lexico-semanticques I*. Presses de l'Univerité de Montréal, Montreal, Canada, 1984.
- [Shell and Carbonell, 1986] Peter Shell and Jaime Carbonell. Frulekit: A Frame-Based Production System. Center for Machine Translation Technical Report, Pittsburgh, PA, 1986.