Submitted to AAAI-96

Technological and Conceptual Tools for Lexical Knowledge Acquisition

Sergei Nirenburg
Kavi Mahesh
Evelyne Viegas
Stephen Beale
Victor Raskin
Boyan Onyshkevych

Computing Research Laboratory
Box 30001
New Mexico State University
Las Cruces, New Mexico 88003
{sergei,mahesh,viegas,sb,raskin,boyan}@crl.nmsu.edu

Abstract

The paper deals with the acquisition of static knowledge sources (ontology and the lexicons) for an NLP system. Acquiring the ontology and lexicon together enables immediate feedback between the two acquisition teams, involving shared semi-automatic acquisition tools. The acquisition tools can be technological (mostly corpus-processing and interface-related) and conceptual (such as the use of lexical rules instead of sense enumeration). We describe how both kinds of tools are used in our approach and present an overall knowledge acquisition methodology. We hope to demonstrate that knowledge acquisition for NLP may be more feasible than in other areas of AI.

Content Areas: knowledge acquisition, ontology, computational lexicography

1. Knowledge Acquisition

Sophisticated environments for knowledge acquisition are becoming increasingly commonplace. When used for practical applications, such tools are typically centered around a user interface. In NLP applications, the interfaces allow for processing text corpora, machine-readable dictionaries and other external resources, as well as provide (often, appropriately structured) editing environments for acquisition of lexicons, grammars and other knowledge bases. A different class of acquisition tools – which can be called conceptual – helps automate a part of the acquisition process by augmenting the knowledge acquired by the human acquirer through the use of special acquisition rules. A number of such tools can be used together in a single environment. To facilitate such integration, a general acquisition methodology must be formulated. This latter task can only be effectively carried out in NLP applications once the domain of knowledge, the languages involved and the type of processing are determined. The methodology then becomes a *situated* methodology. In this paper we attempt to illustrate one such methodology, and one class each of the interface and conceptual tools.

We situated our acquisition in the area of knowledge-based machine translation (KBMT). KBMT views the task of translating a text in one natural language to a different natural language as a problem in mapping meanings. The semantics of the source language, rather than just its syntax, must be mapped to that of the target language. A basic requirement for doing this is a representation of text meaning which is not in the source or target languages. This interlingual meaning representation must be grounded in a language-independent world model, or *ontology* that supplies not only the primitive symbols for the representation but also a set of well-defined composition operations for combining primitive symbols to represent bigger chunks of meaning.

Mikrokosmos (MikroKosmos) is a KBMT system under development at New Mexico State University jointly with the US Department of Defense (Beale, Nirenburg, and Mahesh, 1995; Mahesh and Nirenburg, 1995a, 1995b; Onyshkevych and Nirenburg, 1994). Unlike many previous projects in interlingual MT, MikroKosmos is a large-scale, practical MT system focusing currently on translating Spanish news articles to English. By the end of 1995, a lexicon of approximately 7,000 Spanish words supported by an ontology of about 5,000 concepts was in place. High quality semantic analyses of up to 10 article-length Spanish texts in the domain of company mergers and acquisitions have already been produced.

Figure 1 illustrates the interactive knowledge acquisition tools and environment. Note that in our approach the development of the ontology and the lexicons is carried out simultaneously with the development and testing of the processing components of the system (represented int he figure by the analyzer). As can be seen from the density of the links among the human roles in the process, acquisition is highly interdependent. In Section 2 we describe our methodology and interactive tools for ontology acquisition. In section 3 we describe the conceptual tools on the example of the lexical rule mechanism used in the acquisition of the lexicon.

2. Interactive Tools for Ontology Acquisition

The ontology provides concepts for representing word meanings in the lexicons for the source or target languages. The figure shows an analysis lexicon where words are mapped to concepts in the ontology along with any modifiers that augment the selectional constraints represented by the conceptual relationships in the ontology. Such modification enables the system to capture the many nuances of meanings in different languages. Every lexicon in the MikroKosmos system, irrespective of its language and whether it is constructed for analysis or for generation, maps to the same set of concepts in the ontology.

A *situated* ontology, then, is a world model used as a computational resource for solving a particular set of problems (Mahesh and Nirenburg, 1995a). It is treated as neither a "natural" entity waiting to be discovered nor a purely theoretical construct. World models (ontologies) in computational applications are artificially constructed entities. They are created, not discovered. Many ontologies are developed for purely theoretical purposes and never really constructed to become a computational resource. Even those that are constructed, Cyc (Lenat and Guha, 1990) being the best example, are often developed

Figure 1: The Interactive Tools and Acquisition Environment in Mikrokosmos

without the context of a practical situation (e.g., Smith, 1993). Many practical knowledge-based systems, on the other hand, employ world or domain models without recognizing them as a separate knowledge source (e.g., Farwell, et al. 1993). In the field of NLP, there is now a consensus that all NLP systems that seek to represent and manipulate meanings of texts need an ontology (e.g., Bateman, 1993; Nirenburg, Raskin, and Onyshkevych, 1995). In our continued efforts to build a multilingual KBMT system using an interlingual meaning representation (e.g., Onyshkevych and Nirenburg, 1994), we have developed an ontology to facilitate natural language interpretation and generation.

Not only is the MikroKosmos ontology situated in our machine translation architecture, its development is also very much situated in the complex processes of lexical knowledge acquisition, development of analysis programs, and system testing and evaluation (see Figure 1). The primary source of meanings to be encoded as new concepts in the ontology is the continual stream of requests for concepts from lexicographers trying to represent meanings of words using concepts in the ontology. Concepts acquired per such requests are in turn tested in the semantic analyzer almost immediately and changes and corrections sent to the ontology acquirers within a few days.

Moreover, since the number of people browsing the ontology (lexicographers, system builders, and testing and evaluation experts, over 10 in our situation) is many times the number of ontology developers (at most two in our case), it is highly likely that any error will be noticed during browsing, especially by those who requested the concept in error, and corrected through the feedback process. In fact, because of this imbalance in the numbers of ontology developers and customers, we had to resort to computer support in the form of interfaces and communication and bookkeeping programs to assist in keeping track of requests and complaints to ontology developers. Ontology developers at present interact regularly with lexicographers building lexicons in Spanish, Japanese, and Russian.[1]

Ontology acquisition involves deciding what knowledge to acquire, formulating the knowledge according to the principles and guidelines behind the ontology, and then actually representing the knowledge according to the structure and semantics of the representation language. In contrast to the methods employed by practically every other group, we have chosen a situated approach, where we are driven by the immediate needs of our entire development environment, not just one lexicon or one task such as semantic analysis, as all the static (lexicons, grammars, ontologies) and dynamic (parsers, semantic analyzers, etc.) knoweldge sources are acquired simultaneously.

This negotiation to meet the constraints on both a lexical entry and a concept in the ontology leads to the best choice in each case. It also ensures that every entry in each knowledge base is consistent, compatible with its counterparts, and has a purpose towards the ultimate objective of the task such as producing quality TMRs. The ideal method for situated development of knowledge sources for multilingual NLP is one where an ontology and at least two lexicons for different languages are developed concurrently. This ensures

[1]Construction of Japanese and Russian lexicons has begun recently. Although the ontology is already being used for representing lexical meanings in all three languages, the MikroKosmos analyzer has so far only been tested on Spanish texts.

that the ontology is truly language independent and that representational needs of more than one language are taken into account.

The concepts we acquire are also immediately put to test in more than one component of the project giving us immediate feedback on their correctness and usefulness. Decisions in lexical semantics and concept acquisition are often made independently by different people with different constraints in mind. In spite of significant differences in constraints, goals, and intuitions, the situated methodology affords a protocol and an environment for fruitful negotiations to achieve total compatibility between lexical and ontological knowledge representations.

Ontology acquisition is a very expensive empirical task. Situated development is a good way to constrain the process and make it attainable. For example, in the NLP situation, the acquirer must focus on concepts in the domain of the input texts (i.e., a corpus) and thereby increase the ratio of the number of concepts (and their PROPERTYs) that are actually used in processing a set of texts to the total number of concepts (and their PROPERTYs) encoded in the ontology. The best example of a large ontological database acquired with enormous efforts but entirely outside the context of a particular task is CYC (Lenat and Guha, 1990). While the utility of CYC in a particular situation such as large scale NLP is yet to be demonstrated, it is also true that most projects cannot afford to spend as many resources as it has taken to develop CYC and must strive to constrain acquisition significantly or share existing ontologies.

2.1. Technology for Ontology Development

In order to aid ontology acquisition and maintenance, to check its consistency, to monitor its quality, and to support interactions with lexicographers, a variety of semi-automated tools have been developed and deployed in the MikroKosmos project (see Figure 1). Tools are in use for:

- browsing the ontological hierarchies and the internals of concepts in the ontology;
- graphical editing support: the "Mikrokarat" tool[2] supports complete functionality for editing the graph structures in an ontology;
- translating between two different representations: the object-oriented one suitable for computational purposes and the plain text representation that is more suitable for certain other programs and manual search and maintenance purposes;
- various types of consistency checking both within the ontology and with lexicons, and for conformance with the guidelines;
- supporting interactions with lexicon acquirers through an interface for submitting requests for changes or additions (see below).

[2]Developed by Ralf Brown at the Center for Machine Translation, Carnegie Mellon University.

This set of tools is being shared across geographical, disciplinary, and project group boundaries on a daily basis. For example, on a typical day, 8-10 people browse the MikroKosmos ontology and one or two people develop it using the Mikrokarat tool.

3. Lexical Rules in Lexicon Acquisition

In this section, we will illustrate our general methodology of semi-automatic lexical acquisition and focus on the role of lexical rules as a conceptual tool to complement the interface-oriented tools used by the acquirers. The latter include all the necessary capabilities for accessing corpora, machine-readable dictionaries, lexicons and the ontology. Additionally, there is an entry acquisition scenario consisting of a series of predefined semantico-syntactic templates, which guide the acquirer through lexicon entry acquisition.

In our work, we use the interactive tools to acquire directly about 20% of the lexicon. The balance of the entries is acquired from this 20% automatically, with the help of lexical rules. In this paper we concentrate on the latter.

In this section, we discuss the discovery, representation, and use of lexical rules (LRs). LRs are viewed as a means to minimize the need for costly lexicographic heuristics and to reduce the number of lexicon entry types which must be acquired interactively. The findings reported here have been implemented and tested on the basis of Spanish and English business- and finance-related corpora.

The central idea of our approach to LRs – that there are systematic paradigmatic meaning relations between lexical items, such that, given an entry for one such item, other entries can be derived automatically – is certainly not novel. In modern times, it has been (re)introduced into linguistic discourse by the Meaning-Text group in their work on lexical functions (see, for instance, Mel'čuk 1979). It has been lately incorporated into computational lexicography in Atkins (1991), Ostler and Atkins (1992), Briscoe and Copestake (1991), Copestake and Briscoe (1992), Briscoe et al. (1993).

Pustejovsky (1991, 1995) has coined an attractive term to capture these phenomena: his 'generative lexicon' is promoted as a departure from sense enumeration to sense derivation with the help of lexical rules. The generative lexicon provides a useful framework for potentially infinite sense modulation in specific contexts (cf. Leech 1981, Cruse 1986), due to type coercion (e.g., Pustejovsky 1993) and similar phenomena. Most LRs in the generative lexicon approach, however, deal with small classes of words and explain such grammatical and semantic shifts as $+count$ to $-count$ or $-common$ to $+common$.

While shifts and modulations are important, we find that the main significance of LRs is their promise to aid the task of *massive* lexical acquisition. We expect that in our approach each LR will drastically minimize the effort and expense of acquisition of hundreds, if not thousands, of lexical entries.

Section 3.1 outlines the nature of LRs in our approach and their status in the computational process. Section 3.2 presents a fully implemented case study, the morpho-semantic

LRs. Section 3.3 briefly reviews the cost factors associated with LRs.

3.1. Nature of Lexical Rules

The basic unit of the lexicon is a 'superentry,' one for each citation form, irrespective of its lexical class. Word senses are called 'entries.' The LR processor applies to all the word senses for a given superentry. For example, *pronunciar* has (at least) two entries (one could be translated as "articulate" and one as "declare"); the LR generator, when applied to the superentry, would produce (among others) two forms of *pronunciacion*, derived from each of those two senses/entries.

There need not be any correlation between syntactic category and semantic or ontological class. For example, although meanings of many verbs are represented through reference to ontological EVENTs and a number of nouns are represented by concepts from the OBJECT sublattice, frequently this is not the case. Many LRs change the syntactic category of the input form; in our model the semantic category is preserved in many of these LRs. For example, the verb *destroy* may be represented by an EVENT, as will the noun *destruction* (naturally, with a different linking in the syntax-semantics interface). Similarly, *destroyer* (as a person) would be represented using the same event with the addition of a HUMAN as a filler of the agent case-role. This built-in transcategoriality strongly facilitates applications such as interlingual MT, as it renders vacuous many problems connected with category mismatches and misalignments that plague those paradigms in MT which do not rely on extracting language-neutral text meaning representations. This transcategoriality is supported by LRs.

Once LRs are defined in a computational scenario, a decision is required on the time of application of those rules. A hybrid scenario is also plausible, where, for example, LRs apply at acquisition time to produce new lexical entries, but may also be available at run time as an error recovery strategy to attempt generation of a form or wordsense not already found in the lexicon.

- Acquisition Time - LRs can be used at the time the lexicographers are creating the lexicon to suggest fully-formed additional lexical entries derived from the manually-built base form. The major advantage of this strategy is that the results of any LR expansion can be checked by the lexicon acquirer, though at the cost of substantial additional time. Even with the best left-hand side (LHS) condition (see below), the lexicon acquirer may be flooded by new lexical entries to validate. During the review process, the lexicographer can accept the generated form, reject it as inappropriate, or make minor modifications. If the LR is being used to build the lexicon up from scratch, then mechanisms used by Ostler and Atkins (1992) or Briscoe (1995), such as blocking or preemption, are not available as automatic mechanisms for avoiding overgeneration. Those mechanisms assume that the entire lexicon is in place other than the words being generated; in a lexicon under construction, there are no candidates to block or preempt overgenerated forms.

- Lexicon Load Time - The LRs can be applied to the base lexicon at the time the lexicon is loaded into the computational system. As with run-time loading, the risk is that overgeneration will cause more degradation in accuracy than the missing (derived) forms if the LRs were not applied in the first place. If the LR inventory approach is used or if the LHS constraints are very good (see below), then the overgeneration penalty is minimized, and the advantage of a large run-time lexicon is combined with efficiency in look-up and disk savings.

- Run Time - Application of LRs at run time raises additional difficulties by not supporting an index of all the head forms to be used by the syntactic and semantic processes. For example, if there is an LR which produces *abusive-adj2* from *abuse-v1*, the adjectival form will be unknown to the syntactic parser, and presumably its production would be triggered by failure recovery mechanisms if direct lookup failed and the reverse morphological process identified *abuse-v1* as a potential source of the entry needed.

For any of the LR application opportunities itemized above, the methodology needs to account for the selection of specific LRs as applicable to a given lexical entry (whether base or derived). Without one of these two approaches (or, again, a combination), the LRs will grossly overgenerate, resulting in inappropriate entries, computational inefficiency, and degradation of accuracy.

- LR Itemization - The simplest mechanism of rule triggering is to include in each lexicon entry an explicit list of applicable rules. LR application can be chained, so that the rule chains must be expanded, either statically, in the specification, or dynamically, at application time. This approach avoids any inappropriate application of the rules (overgeneration), though at the expense of tedious work at lexicon acquisition time. One drawback of this strategy is that if a new LR is added, each lexical entry needs to be revisited and possibly updated.
- Rule LHS Constraints - The other approach is to maintain a bank of LRs, and rely on the LHSs to constrain the application of the rules to only the appropriate cases; in practice, however, it is difficult to set up the constraints in such a way as to avoid over- or undergeneration *a priori*. Additionally, this approach (at least, when applied after acquisition time) does not allow explicit ordering of word senses, a practice preferred by many lexicographers to indicate relative frequency or salience; this sort of information can be captured by other mechanisms (e.g., using frequency-of-occurrence statistics). This approach does, however, capture the paradigmatic generalization that is represented by the rule, and simplifies lexical acquisition.

3.2. The Cost of Lexical Rules: The Case of Derivational Morphology

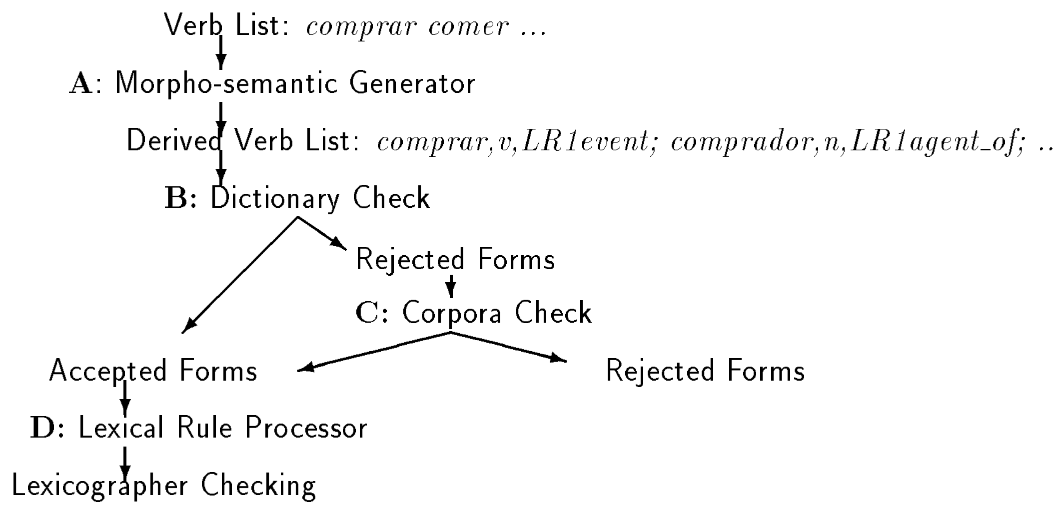
It is clear by now that LRs are most useful in large-scale acquisition. It should be made equally clear, however, that the use of LRs is not cost-free. We illustrate our point by discussing LRs based on constructive derivational morphology. Such LRs automatically derive word forms which are polysemous, such as the Spanish *generador* 'generator,' either the artifact or someone who generates. The LRs have been tested in a real world application, involving the semi-automatic acquisition of a Spanish computational lexicon of about 7,000 different word meanings.

An LR-relevant fragment of the Spanish lexicon entry for *comprar*, 'buy,' below illustrates the knowledge we acquire and the format in which we encode it:

```
(comprar
   (comprar-V1                            ;the first verbal sense of comprar
      (CAT v)
      (ANNO
         (DEF "acquire the possession or right by paying or promising to pay")
         (EX "Mi papa compro un coche nuevo ayer")
         (COMMENTS Trs-Ag-Th)
         (TIME-STAMP "18/1 15:42:44")
         (SYN-STRUC                       ;syntactic characteristics:
           (1 ((root $var0)               ;subcategorisation pattern
                (cat v)                   ;$var0 is bound to the verb itself
                (subj ((root $var1)       ;$var1 is bound to an np
                        (cat np)))        ;the np is the grammatical subject
                (obj ((root $var2)        ;$var2 is bound to an np
                       (cat np))))))      ;the np is the grammatical object
         (SEM-STRUC                       ;semantic information
           (LEX-MAP                       ;the syntax-semantics mapping
            (1  (buy                      ;ontological concept
                (agent (value ^$var1))    ;case-role for the meaning of ^$var1
                (theme (value ^$var2)))))) ;case-role for the meaning of ^$var2
                ...
)
```

Typically, a lexicographer types in the definition (`DEF`) and the example (`EX`). The `SYN-STRUC` and `SEM-STRUC` are acquired by selecting one of a set of predifined templates provided by the interactive acquisition tool. The template **Trs-Ag-Th** specifies that this verb subcategorises for two arguments, namely a subject and an object, whose semantic roles are agent and theme, respectively. BUY is an ontological concept, which contains properties and constraints specifying selection restrictions on the agent and theme, HUMAN and OBJECT, respectively. We accelerated the process of acquisition by developing morpho-semantic LRs which, when applied to a lexeme, produced an average of 25 new candidate entries. Our experience shows that it is much easier for a lexicographer tocheck an entry than to create it from scratch, which typically requires special skills and experience. These rules provide an efficient and cheap way to expand the size of a lexicon.

The diagram below illustrates the overall process of generating new entries from a citation form, by applying morpho-semantic LRs.



Generation of new entries usually starts with verbs. Each verb found in the corpora, is submitted to the morpho-semantic generator which produces all its morphological derivations and, based on a detailed set of tested heuristics, attaches to each form an appropriate semantic LR label, for instance, the nominal form *comprador* will be among the ones generated from the verb *comprar* and the semantic LR "agent-of" is attached to it, as seen below, where we present a partial list of the morphology derivations and associated semantics for *comprar*:

comprar, v, LR1event
comprador, n, LR2agent_of1a
compra, n, LR2event10
compra, n, LR2theme_of_event10
compradero, adj, LR3feasibility_attribute2a
comprable, adj, LR3feasibility_attribute1
comprado, adj, LR3event_telic
compradizo, adj, LR3feasibility_attribute5a
comprador, adj, LR3social_role_relation1a
malcomprar, v, LRneg_affect1, LR1event
malcomprado, adj, LR3event_telic
recomprar, v, LRrepetition1, LR1event
recompra, n, LR2event10
recompra, n, LR2theme_of_event10
recomprado, adj, LR3event_telic

The mechanism of rule application is illustrated below.

The Lexical Rule Processor is an engine which produces a new entry from an existing one, such as the new entry *compra* produced from the verb entry *comprar* shown earlier after applying the LR2event rule:

```
(compra
        (compra-N1
              (CAT n)
              (ANNO
                     (DEF "acquire the possession or right by paying
                          or promising to pay")
                     (EX " ")
                     (COMMENTS Comun  )
                     (TIME-STAMP "LR2event","11/12 14:01:12"))
              (SYN-STRUC
                     (1 ((root $var0)
                         (cat n))))
              (SEM-STRUC
                 (LEX-MAP
                     (1 (buy))))
              (LEX-RULES comprar-V1 "LR2event")
              ...
))
```

The acquirer must check the definition and enter an example, but the rest of the information is simply retained. The LEXical-RULES zone specifies the morpho-semantic rule which was applied to produce this new entry and the verb it has been applied to.

Besides the effort of discovering and implementing them, there is also the significant time and effort expenditure on the procedure of semi-automatic checking of the results of the application of LRs to the basic entries, such as those for the verbs.

The shifts and modulations studied in the literature in connection with the LRs and generative lexicon have also been shown to be not problem-free: sometimes the generation processes are blocked–or preempted–for a variety of lexical, semantic and other reasons (see Ostler and Atkins 1992). In fact, the study of blocking processes, their view as systemic rather than just a bunch of exceptions, is by itself an interesting enterprise (see Briscoe et al. 1995).

Obviously, similar problems occur in real-life large-scale lexical rules as well. Even the most seemingly regular processes do not typically go through in 100 % of all cases. This makes the LR-affected entries not generable fully automatically and this is why each application of a LR to a qualifying phenomenon must be checked manually in the process of acquisition.

We have discussed several aspects of the discovery, representation, and implementation of LRs, where, we believe, they count, namely, in the actual process of developing a real-life NLP system. Our LRs tend to be large-scope rules, which saves us a lot of time and effort on massive lexical acquisition. The LRs are certainly most useful right there: they make it unnecessary to figure out the representation of many lexical entries by deriving them instead from those already represented earlier. But are LRs as uniformly good as the literature often presents them? Do they help with everything we do?

It is clearly very exciting to discover rules and regularities in the lexicon, where a mere enumeration reigned supreme–or so some thought–until very recently. It is naive, however, to think, as Pustejovsky strongly implies in his discussions of the generative lexicon, that no senses need to be enumerated at all. This is not our experience: we find that if a lexical entry has several meanings, they need to be enumerated, and LRs may, in fact, be an efficient tool for enumerating them. The alternative, i.e., trying to derive all of them from one meaning, is burdensome and often impractical, even if feasible. The generative lexicon seems to be a good approach for contextual meaning shifts and modulations, and we do not deny the importance of those. We claim, however, that such shifts and modulations do not occur as frequently as the simpler type of compositionality, which is basically found in every clause, and our emphasis is on making the computation of "ordinary" sentences optimal with the help of the entries, many, if not most, of which are produced with the help of LRs.

It is in this context that we have had to examine when to use LRs, at what stage of the process to trigger them, and how to represent them in the basic lexical entries. While the literature on LRs can be seen as hailing the postulation of as many LRs as can be found, even if their scope is very limited, we have to be careful with overburdening the lexicon and analyzer with LRs which do not deliver too many goods to us. Each LR comes at a certain human-labor and computational expense, and if their pay load is very limited, they may not be worth the extra effort.

We cannot say at this point that LRs provide any advantages in computation or quality of the deliverables. What we do know is that, when used justifiably and maintained at a large scope, they facilitate tremendously the costly but unavoidable process of semi-automatic lexical acquisition.

4. Conclusion

In this paper, we have argued for a situated methodology for the acquisition of both the ontology and the lexicons based on it. Under such a methodology, the ontological and lexical acquisition are tightly interwoven with the help of multiple mutual-feedback loops; they share a set of semi-automatic acquisition tools. We also argue strongly for the use of large-scope ontologically-based lexical rules, which expedite considerably the process of massive lexical acquisition, establishing in the process the transcategoriality of lexical entries based on the same ontological concept. We hope that this paper makes it reasonably obvious that large-scale knowledge acquisition for NLP may be much more feasible–and, in fact, simpler–than in other areas of AI.

Acknowledgements

References

Apresjan, Ju.D. 1976. Regular Polysemy. *Linguistics* vol 142, pp. 5-32.

Atkins, B. T. S. 1991. Building a lexicon:The contribution of lexicography. In B. Boguraev (ed.), "Building a Lexicon", Special Issue, *International Journal of Lexicography* 4:3, pp. 167-204.

Boguraev, B. and J. Pustejovsky. 1990. *Knowledge Representation and Acquisition from Dictionary.* Coling Tutorial, August 16-18, 1990, Helsinki, Finland.

Briscoe, E.J. and A. Copestake. 1991. Sense extensions as lexical rules. In *Proceedings of the IJCAI Workshop on Computational Approaches to Non-Literal Language.* Sydney, Australia, pp. 12-20.

Briscoe, E.J., Valeria de Paiva, and Ann Copestake (eds.). 1993. **Inheritance, Defaults, and the Lexicon**. Cambridge: Cambridge University Press.

Briscoe, E.J., Ann Copestake, and Alex Lascarides. 1995. Blocking. In P. Saint-Dizier and E.Viegas (eds.), **Computational Lexical Semantics**. Cambridge University Press.

Copestake, A. and T. Briscoe. 1992. Lexical operations in a unification-based framework. In J. Pustejovsky and S. Bergler (eds), *Lexical Semantics and Knowledge Representation.* Berlin: Springer, pp. 101-119.

Cruse, D.A. 1986. **Lexical Semantics**. Cambridge: Cambridge University Press.

Leech, G. 1981. **Semantics**. Cambridge: Cambridge University Press.

Mahesh, K. (in preparation) *Ontology Development: Ideology and Methodology.* Computing Research Laboratory, New Mexico State University.

Mel'čuk, I. 1979. **Studies in Dependency Syntax.** Ann Arbor, MI: Karoma.

Mel'čuk, I., N. Arbatchewsky-Jumarie, L. Elnitsky, L. Iordanskaja et A. Lessard (1984) : *Dictionnaire explicatif et combinatoire du fran*{ccais contemporain: recherches lexico-sémantiques I. Montréal: Presses de l'Université de Montréal.

Meyer, I., Onyshkevych, B. et L. Carlson. 1990. *Lexicographic Principles and Design for Knowledge-based Machine Translation.* Technical Report CMU-CMT-90-118, Carnegie Mellon University.

Nirenburg, S., V. Raskin, et B. Onyshkevich. 1994. *Apologiae ontologiae.* Memoranda in Computer and Cognitive Science MCCS-95-281. New Mexico State University: Computing Research Laboratory.

Onyshkevych, B. et S. Nirenburg. 1994. *The Lexicon in the Scheme of KBMT Things.* Technical Report MCCS-94-277, Computing Research Laboratory, New Mexico State University.

Ostler, N and B. T. S. Atkins. 1992. Predictable meaning shift: Some linguistic properties of lexical implication rules. In J. Pustejovsky and S. Bergler (eds), **Lexical Semantics and Knowledge Representation**. Berlin: Springer, pp. 87-100.

Pustejovsky, J. 1991. The generative lexicon. *Computational Linguistics* 17:4, pp. 409-441.

Pustejovsky, J. 1993. Type coercion and lexical selection. In James Pustejovsky (ed.), **Semantics and the Lexicon**. Dordrecht-Boston: Kluwer, pp. 73-94.

Pustejovsky, J. 1995. **The Generative Lexicon.** Cambridge, MA: MIT Press.

Viegas, E. 1995 *Lexicon Development: Pro a Semi-automated Approach.* Technical Report (in preparation). Computing Research Laboratory, New Mexico State University.

Viegas, E. et M. Gonzales 1995. *Derivational Morphology Rules to Enhance Lexicons Acquisition.* Technical Report (in preparation). Computing Research Laboratory, New Mexico State University.

Viegas, E. et S. Nirenburg. 1995. The Semantic Recovery of Event Ellipsis: its Computational Treatment. Proceedings of the Workshop on Context and Natural Language, *IJCAI 95*, Montréal, 1995.